SSAPOV Dictionary

Reference Manual for
Harmonizing Household Surveys in Sub-Saharan Africa

The World Bank

1818 H Street, N.W.

Washington, D.C.

20433, USA.

January 28, 2020

Contents

1 General harmonization guidelines

As Sub-Saharan African economies become more open and globalized, huge opportunities are created for individuals and families. However, a large fraction of households has not benefited sufficiently, and economic and social inequality are real and even growing in some cases. Household surveys provide rich information on living standards and the impact of economic changes on individuals and households. Unfortunately, these data are largely underutilized due to the complexity of household surveys and the significant time required to prepare the survey data for analytical work.

The Sub-Saharan Team for Statistical Development (SSATSD) seeks to eliminate the bottleneck of analyzing household survey data by extracting variables from existing household surveys and ensuring that have the same definition and variable names. These variables include household consumption, access to infrastructure (water, electricity, etc.), employment status, education, and health. Invariably, in each survey, questions will be asked in a different manner, which poses challenges to consistently define harmonized variables. The harmonized household survey data present the best available variables with harmonized definitions.

This document presents detailed guidelines for harmonizing household survey data into a set of commonly defined variables that are available in most types of household surveys. To ensure the quality and transparency of the final harmonized data, it is critical to document the harmonization process and check the final data for quality concerns. This approach assures that the results can be replicated from the original household survey data with ease and that the final data provides reliable temporal and cross-country comparisons.

Four harmonized modules are prepared for each survey. Each of these modules contain a theme of harmonized variables that have the same variable names and definitions. The four harmonized modules are:

1. **Module P: Poverty-related variables**: This module contains consumption variables, regional identifiers, spatial/temporal prices indices, variables indicating national poverty lines, and variables indicating whether households are classified as poor.
2. **Module H: Household-level variables** (except for poverty-related variables): This module contains information on housing amenities, ownership of assets, access to infrastructure and services, and household remittances.
3. **Module I: Individual-level variables** (except labor force variables): This module contains basic characteristics of individuals such as age, sex, literacy, education, and migration status.
4. **Module L: Labor force variables**: This module contains information on labor force variables, such as labor force status, industry, sector of employment, wages, etc.

1.1 datalibweb

In order to ensure the transparency and replicability of the harmonized data, a strict method or organizing folders and files is used. This ensures that different versions of harmonization are kept track of and that users and future revisions of harmonization can be conducted without changing file paths. The method for directory organization and file name conventions follows a practice adopted across regions and implemented through datalibweb. Datalibweb is a data system specifically designed to enable users to access the most up to date versions of non-harmonized (original/raw) and harmonized datasets of different collections across Global Practices. It can easily perform computations relevant for poverty and shared prosperity analysis based on the micro data from different harmonized collections: EAPPOV, ECAPOV, etc. Datalibweb can be installed in two different ways:

1. Directly from Stata: In order to get install to Datalibweb command in Stata, type the following code in the command line, and click on the datalibweb (hyperlink) to install in your computer. "net from http://eca/povdata/datalibweb/_ado"
2. Manual installation: In addition, users can install the package manually. Get the file from this link: http://eca/povdata/datalibweb/_ado/datalibweb.zip. Copy with replacement all the files into c:/ado, without changing the folder structure.

Once datalibweb is installed, and access to data has been granted, all raw data for a survey can be accessed with the following command:

```
datalibweb, country(CCC) year(YYY) type(SSARAW) surveyid(SURVEYNAME) clear
```

where CCC stands for ISO 3 letter country code (see Annex III), YYYY is the survey year according to IHSN standards, which is when the fieldwork started, and SURVEYNAME is the survey acronym.

You should always load data through datalibweb. This assures that no local file paths are used to load the data, and thus that others who have access to the raw data can run the .do-files. All documents related to a survey, such as questionnaires and technical reports, can be accessed through the following command:

```
datalibweb, country(CCC) year(YYY) type(SSARAW) surveyid(SURVEYNAME) request(doc)
```

Once a harmonization is done, the final harmonized files will be stored in datalibweb and can be accessed through the following command:

```
datalibweb, country(CCC) year(YYY) type(SSARAW) surveyid(SURVEYNAME) mod(MODULENAME)
```

where MODULENAME takes the value, P, H, I, or L.

1.2 Folder and File Structure

The back-end of datalibweb contains a very specific folder structure and naming convention. Although we do not work in these folders directly when working with data, it is useful for you to copy the folder structure locally. Before harmonizing a survey, you should first create sub-directories as instructed below. Additionally, all harmonization files must be named per this manual. This rigorous procedure ensures a seamless integration with datalibweb and that different versions of the harmonization are kept track of.

You will get assigned a folder on a server, \\WBGMSAFR1001\AFR_Database\SSAPOV-Harmonization, with his/her name. This should be the parent directory from which all harmonization are saved and from which all work is conducted. This folder should contain subfolders with the ISO3 country codes of the countries with which you is working. Within each country-folder, there should be a folder with the name CCC_YYYY_SURVEYNAME for each of the surveys you has been working on. For example, if a person is working on harmonizing the 2015 HICES survey of Ethiopia, then all material related to this should be saved in this path: \\WBGMSAFR1001\AFR_Database\SSAPOV-Harmonization\[Name]\ETH\ETH_2015_HICES\. This folder should also be the saved as a global in the beginning of each .do-file.

Each survey-specific folder should have two subfolders with the following content:

- Programs: This folder should contain only the 4 .do-files used to construct the different modules. If some preliminary data cleaning is needed, this should be included in the other .do-files. The .do-files should not call each other or any other .do-files.
- Data\Harmonized: This folder should contain the 4 .dta-files with the harmonized modules

The .do-files .dta files related to harmonization should be named according to the following convention:

CCC_YYY_SURVEYNAME_v0x_M_v0y_A_SSAPOV_MODULENAME.do

CCC_YYY_SURVEYNAME_v0x_M_v0y_A_SSAPOV_MODULENAME.dta

Here "v0x" is the version of the raw data. This will almost always be v01, but if errors were found in the original data and a new version of data is received from the National Statistical Office, then it will be called v02, etc. "v0y" is the version of the harmonized data. This will often be v01, but if revisions are made and the .do-file needs to be updated then the new .do and .dta files will be named v02, etc. This assures that anyone can run the .do file without any changes and code and that, if the path becomes outdated, only one line of code needs to be changed.

1.3 Guidelines across Modules

The following harmonization guidelines apply to all modules:

- To the extent possible, all variables should be arranged in the order they appear in this manual.
- Frequently, surveys do not have information on all variables that we seek to harmonize. In this case, the variables should still be created as missing such that all variables appear in all modules.
- In the P and H-modules the household identifier hid must uniquely identifies observations. That is the isid hid command should not return an error. Likewise, it is important in the I and L-modules that hid and pid uniquely identify observations. That is isid hid pid should not return an error. An implication of this is that hid (and pid in the I and L-modules) should have no missing values.
- The same households do not need to appear in all four modules. There may be some households in the H, I, and L modules but may not appear in the P module if the household did not respond to the consumption module. However, if a household is present in the P module but not in the H module, this may mean that the household id may have been miscoded in either the H or the P module. In general, we want to keep all households used for computing the poverty rate at the national poverty line.
- Any critical assumption that is made during the harmonization process should be stated clearly in the .do-file under a comment heading.
- The labels for all variables should be created at the end of each .do file. This may involve creating new variables that are a function of other harmonized variables.

1.4 Missing Value Codes

You should differentiate missing values of variables from variables that were present in the survey but could not be harmonized due to time constraints. This will help the others to focus on the unharmonized variables. The missing value code for these two scenarios are:

- For variables unavailable in survey = .
- For variables available in the survey but not harmonized = .a . To do so use this Stata command: `gen str varname = .a` if the variable is a string and `gen double varname= .a` if the variable is numeric.

1.5 Qcheck

Once a module is harmonized, a quality check will be performed on the harmonized data using a program called qcheck. Qcheck tests if all variables are in the dataset, if all variables have the correct format, if the variables take plausible values, and if some of the variables are mutually inconsistent. For example, the age variable may be negative, which would indicate an error. It will also flag if someone is coded to have no education in one education variable but have completed secondary education in another variable.

2 P Module – Poverty-related Variables

The most common measures used for living standards are consumption and income. Income refers to actual earnings from productive activities and transfers, while consumption refers to resources consumed. While income may be used as an indicator to measure welfare, it is not ideal in countries where much of the population works in informal sectors, such as small business, work on land, etc., as net income becomes very difficult to measure in these cases. Additionally, incomes may be zero or negative for self-employed workers during a given timeframe, even though these individuals could have wealth to draw upon. In these cases, income is a poor proxy for welfare. Consumption is therefore thought to provide a better picture of a household's standard of living than a measure of current income.

For these reasons, most countries in Sub-Saharan Africa use consumption to measure poverty. The p-module contains a list of variables related to consumption, such as its breakdown by food and non-food consumption, consumption per capita and per adult equivalent, as well as indicators for whether a household's consumption falls short of the poverty line.

There are limitations of household surveys in measuring household consumption:

- A household survey relies mostly on self-reported data and on household members' memory. The latter makes estimates heavily dependent on the length of the recall period.
- It is practically impossible to distinguish between consumption and monetary expenditures. What was bought may also not necessarily be consumed by households in its entirety and thus it becomes difficult to separate consumption and expenditure.
- The recall period may lead to either underestimation or overestimation of the reported data, and thus expenditure consumption surveys should be designed to envisage such a problem.
- A perennial issue relating to national income in any country has been the difference between the System of National Accounts (SNA) Statistics and National Sample Survey estimates on consumption expenditure. The SNA private household consumption expenditure is available as an estimate for the entire nation, while the National Sample Survey consumption estimates are available for sub-groups such as provinces, rural, and urban areas among others, which can be aggregated to derive a national estimate. The estimates of private consumption from these two sources are different, primarily because of conceptual differences and estimation approaches.
- Consumption aggregates are not comparable across households if prices differ across time and space. For this reason, a lot of effort goes into adjusting the consumption aggregates temporally and spatially. The P-module contains several variables trying to document whether spatial and or temporal deflation was used for a specific survey, both for purposes of national poverty estimation and for purposes of international poverty estimation.

2.1 Sample, Geography, and Basic Household Identifiers

Variable: harmonization
Label: Type of harmonization
Type: String variable
Description: use the following code to generate:
gen harmonization = "SSAPOV"

Variable: country
Label: Country code
Type: String variable
Description: 3-character length (Annex IV)

Variable: survey
Label: Type of survey
Type: String variable
Description: Specifies the type of survey. Possible names are: HBS, LSMS, IS, CWIQ, etc. Upper-case letters should be used.

Variable: survey_coverage
Label: Survey coverage
Type: Numeric categorical variable
Description: 1 = National; 2 = Urban; 3 = Rural; 4 = Other

Variable: usemicrodata
Label: Use of microdata
Type: Numeric categorical variable
Description: 0 = Grouped; 1 = Micro

Variable: year_IHSN
Label: 4-digit year of survey based on IHSN standards
Type: Numeric discrete variable
Description: This is the start year of survey based on the IHSN standards. It should be identical to the year used for file-naming purposes.

Variable: region1
Label: Subnational ID – highest level
Type: String variable
Description: This variable should contain the first-level administrative divisions of a country. It should contain numeric entries in string format using the following naming convention: "1 – Hatay" (as string). The code below shows how to turn a numeric variable with labels into the format required:

```
gen region1=""
qui levelsof inputvar, local(lev)
foreach cc of local lev {
       cap loc la_`cc': label(inputvar) `cc'
       if !_rc    {
              qui replace region1="`cc' - `la_`cc''" if inputvar ==`cc'
       }
}
```

Variable: region2

Label: Subnational ID – second highest level

Type: String variable

Description: This variable should contain the second-level administrative divisions of a country. It should contain numeric entries in string format using the following naming convention: "1 – Hatay" (as string). Use code similar to that for region1 to convert a numeric variable with labels into the format required.

Variable: region3

Label: Subnational ID – third highest level

Type: String variable

Description: This variable should contain the third-level administrative divisions of a country. It should contain numeric entries in string format using the following naming convention: "1 – Hatay" (as string). Use code similar to that for region1 to convert a numeric variable with labels into the format required.

Variable: region4

Label: Subnational ID – fourth highest level

Type: String variable

Description: This variable should contain the fourth-level administrative divisions of a country. It should contain numeric entries in string format using the following naming convention: "1 – Hatay" (as string). Use code similar to that for region1 to convert a numeric variable with labels into the format required.

Variable: subnatidsurvey

Label: Lowest level of subnational ID

Type: String variable

Description: subnatidsurvey is a string variable that refers to the lowest level of the geographic units at which the survey is representative. In most cases this will be equal to "region1" or "region2". It should contain numeric entries in string format using the following naming convention: "1 – Hatay" (as string). Use code similar to that for region1 to convert a numeric variable with labels into the format required. However, in some cases the lowest level is classified in terms of urban, rural or any other regional categorization cannot be mapped to regions. The variable would contain survey representation at lowest level irrespective of its mapping to regions.

Variable: region1_prev

Label: Subnational ID of most recent previous survey

Type: String variable

Description: Variable is coded as missing unless the classification used for region1 has changed since the most recent previous survey.

Variable: region2_prev

Label: Subnational ID of most recent previous survey

Type: String variable

Description: Variable is coded as missing unless the classification used for region2 has changed since the most recent previous survey.

Variable: region3_prev

Label: Subnational ID of most recent previous survey

Type: String variable

Description: Variable is coded as missing unless the classification used for region3 has changed since the most recent previous survey.

Variable: region4_prev

Label: Subnational ID of most recent previous survey
Type: String variable
Description: Variable is coded as missing unless the classification used for region4 has changed since the most recent previous survey.

Variable: strata

Label: Strata
Type: String variable
Description: strata refer to the division of the target population – typically the census sample frame -- into subpopulations based on auxiliary information that is known about the full population. Sampling is conducted separately for each stratum. The strata should be mutually exclusive: every element in the population must be assigned to only one stratum. The strata should also be collectively exhaustive: no population element can be excluded. Sampling strata need to be considered when constructing the variance (or confidence intervals) of population estimates. strata is needed for the correct calculation of standard deviation for each sample design. Strata is numeric and country-specific. A unique identifier is created for each stratum. In STATA, users are advised to specify strata through the svyset command. The variable is in string format with the following naming convention "code of stratum – stratum name", for example: "1 – Dar-es-salaam"

Variable: rururb

Label: Area of residence
Type: Numeric categorical variable
Description: Each country defines this jurisdiction according to a certain criterion. In transition economies where 'semi-urban' is a recognized category which includes 'villages of the town type' this will be collapsed into the 'urban' category unless if the country defines these as rural towns.
0 = Rural
1 = Urban

Variable: capital

Label: Capital/city, other urban, and rural classification
Type: Numeric categorical variable
Description: This is a variable which indicates the location of the household's residence. This information can be created from some combination of the strata, region1, or rural/urban variables. The enumerator's manual or the survey report (if available) may help you identify the capital city and other urban areas.
1= Capital city
2= Other urban areas
3 = Rural

Variable: cluster

Label: Primary sampling unit (enumeration area)
Type: Numeric categorical variable
Description: Primary sampling unit based on country requirements.

Variable: gaul_adm1_code

Label: Gaul code for admin1 level
Type: Numeric discrete variable
Description: gaul_adm1_code is numeric and country-specific based on the GAUL database. It should be taken from the same data in the GAUL database where the geographical area can be identified in the survey based on the name of the location/area. The number of unique values from the region1 and the gaul_adm1_code could be

different or the same. Use the following Stata code to find the unique list of gaul_adm1 codes for your country (in this case, RWA):

```
Use “GAUL codes for SSAPOV harmonization.dta", clear
keep if countrycode=="RWA"
duplicates drop wb_adm1_co wb_adm1_na if countrycode=="RWA", force
li wb_adm1_co wb_adm1_na
```

Variable: gaul_adm2_code

Label: Gaul code for admin2 level
Type: Numeric discrete variable
Description: gaul_adm2_code is numeric and country-specific based on the GAUL database. It should be taken from the same data in the GAUL database where the geographical area can be identified in the survey based on the name of the location/area.

```
Use “GAUL codes for SSAPOV harmonization.dta", clear
keep if countrycode=="RWA"
duplicates drop wb_adm2_co wb_adm2_na if countrycode=="RWA", force
li wb_adm2_co wb_adm2_na
```

Variable: hhno

Label: Household number
Type: Numeric discrete variable
Description: Household number

Variable: hid

Label: Household unique identification
Type: string or numeric, of original data should be kept
Description: This variable should uniquely identify observations and cannot be missing, i.e. isid hid should return no error.

Variable: hid_orig

Label: Household identifier in the raw data
Type: string or numeric, of original data should be kept
Description: This variable is missing if the raw data does not have hid and should be created using other variables (such as region, sector, etc.) . This is the household ID that was included in the raw data.

Variable: int_month

Label: Month of interview visit
Type: Numeric discrete variable
Description: The month when the survey questionnaire was administered to the household. This variable will take on values 1-12, with 1 representing January and 12 representing December.

Variable: int_year

Label: Year of interview visit
Type: Numeric discrete variable
Description: The year when the survey questionnaire was administered to the household.

Variable: hhsize

Label: Household size
Type: Numeric discrete variable
Description: Total number of residents (regular members).
The definition of regular member is country-specific.

Variable: ctry_adq

Label: Adult equivalent scale

Type: Numeric continuous variable

Description: Definition varies from country to country, as different adult scales exist worldwide. Total number of adult equivalent people in household must be greater 0 and less than or equal to hhsize (household size). This variable is usually provided by the NSO.

Variable: wta_hh

Label: Household weights

Type: Numeric continuous variable

Description: To obtain household estimates, this is the weight to be used in all computations referring to household-level estimates. This variable cannot be used for poverty estimation. The interpretation is the proportion of households with a certain characteristic is XX%.

Variable: wta_pop

Label: Population weights

Type: Numeric continuous variable

Description: This variable should be used for poverty estimation. The interpretation is the proportion of individuals with a certain characteristic is XX%.

```
gen wta_pop = wta_hh*hhsize
```

Variable: wta_cadq

Label: Adult equivalent weights

Type: Numeric continuous variable

Description: In a number of countries, this weight is used to derive the proportion of poor population. The interpretation is the proportion of adult equivalent population with a certain characteristic is XX%. `gen wta_cadq = wta_hh* ctry_adq`

2.2 Consumption Expenditure Values

Variable: welfaretype

Label: Type of welfare measure (income, consumption, expenditure)

Type: String variable

Description: Specifies the type of welfare aggregate used for poverty estimation in a country. This variable should equal “CONS”, “INC”, or “EXP”. CONS=consumption; INC=income; EXP=expenditure

Variable: fdtexp

Label: Purchased and auto-consumption food expenditure, nominal (annual)

Type: Numeric continuous variable

Description: Country-derived by the NSO.

Variable: nfdtexp

Label: Purchased & auto-consumption non-food expenditure, nominal (annual)

Type: Numeric continuous variable

Description: Country-derived by the NSO.

Variable: hhtexp

Label: Household food and non-food consumption expenditure, nominal (annual)
Type: Numeric continuous variable
Description: Country-derived by the NSO.
Use this code to generate hhtexp: `gen hhtexp = fdtexp+nfdtexp`
If the raw data does not separate between food and non-food consumption, create this file instead of letting it be created in the labelling file.

Variable: pc_fd

Label: Per capita food consumption expenditure, nominal (annual)
Type: Numeric continuous variable
Description: Country-derived by the NSO.
Use this code to generate pc_fd*:* `gen pc_fd=fdtexp/hhsize`

Variable: pc_hh

Label: Per capita food and non-food consumption, nominal (annual)
Type: Numeric continuous variable
Description: Country-derived by the NSO.
Use this code to generate pc_hh: `gen pc_hh=hhtexp/hhsize`

Variable: padq_fd

Label: Per adult equivalent food consumption expenditure, nominal (annual)
Type: Numeric continuous variable
Description: Country-derived by the NSO.
Use this code to generate padq_fd*:* `gen padq_fd = fdtexp/ctry_adq`

Variable: padq_hh

Label: Per adult equivalent food and non-food consumption, nominal (annual)
Type: Numeric continuous variable
Description: Country-derived by the NSO.
Use this code to generate padq_hh: `gen padq_hh=hhtexp/ctry_adp`

Variable: fdspindex

Label: Food spatial price index
Type: Numeric continuous variable
Description: Country-derived by the NSO.

Variable: nfdspindex

Label: Non-food spatial price index
Type: Numeric continuous variable
Description: Country-derived by the NSO.

Variable: spindex

Label: Spatial price index
Type: Numeric continuous variable
Description: Country-derived by the NSO.

Variable: fdtpindex

Label: Food temporal price index
Type: Numeric continuous variable

Description: Country-derived by the NSO.

Variable: nfdtpindex
Label: Non-food temporal price index
Type: Numeric continuous variable
Description: Country-derived by the NSO.

Variable: tpindex
Label: Temporal price index
Type: Numeric continuous variable
Description: Country-derived by the NSO.

Variable: fdpindex
Label: Spatial/temporal food index
Type: Numeric continuous variable
Description: Country-derived by the NSO.
This variable should never be missing. If no separate food spatial/temporal price index is used, set this equal to pindex.

Variable: nfdpindex
Label: Spatial/temporal non-food index
Type: Numeric continuous variable
Description: Country-derived by the NSO.
This variable should never be missing. If no separate non-food spatial/temporal price index is used, set this equal to sptpindex.

Variable: pindex
Label: Final spatial/temporal price index
Type: Numeric continuous variable
Description: Country-derived by the NSO. This variable should be the one used to derive wel_PPP and wel_abs. Should never be missing. If no temporal/spatial deflation is used, generate a column of 1's.

Variable: fdtexpdr
Label: Purchased and auto-consumption food expenditure, deflated (annual)
Type: Numeric continuous variable
Description: Use this code to generate fdtexpdr: `gen fdtexpdr = fdtexp/fdpindex`

Variable: nfdtexpdr
Label: Purchased & auto-consumption non-food expenditure, deflated (annual)
Type: Numeric continuous variable
Description: Use this code to generate nfdtexpdr: `gen nfdtexpdr = nfdtexp/nfdpindex`

Variable: hhtexpdr
Label: Household food and non-food consumption expenditure, deflated (annual)
Type: Numeric continuous variable
Description: Use this code to generate hhtexpdr: `gen hhtexpdr = hhtexp/pindex`

Variable: pc_fddr
Label: Per capita food consumption expenditure, deflated (annual)
Type: Numeric continuous variable

Description: Use this code to generate pc_fddr: `gen pc_fddr = fdtexpdr/hhsize`

Variable: pc_hhdr
Label: Per capita food and non-food consumption expenditure, deflated (annual)
Type: Numeric continuous variable
Description: Use this code to generate pc_hhdr: `gen pc_hhdr = hhtexpdr/hhsize`

Variable: padq_fddr
Label: Per adult equivalent food consumption expenditure, deflated (annual)
Type: Numeric continuous variable
Description: Use this code to generate padq_fddr: `gen padq_fddr = fdtexpdr/ctry_adq`

Variable: padq_hhdr
Label: Per adult equivalent food & non-food consumption expenditure, deflated (annual)
Type: Numeric continuous variable
Description: Use this code to generate padq_hhdr: `gen padq_hhdr = hhtexpdr/ctry_adq`

Variable: wel_abs_deflation
Label: Spatial/temporal deflation used for national poverty estimation
Type: Numeric categorical variable
Description:
0 = Neither spatially nor temporally deflated
1 = Spatially deflated
2 = Temporally deflated
3 = Both spatially and temporally deflated

Variable: wel_abs_pcpadq
Label: Per adult equivalent or per capita adjustment used for national poverty estimation
Type: Numeric categorical variable
Description:
0 = Per capita
1 = Per adult equivalent

Variable: wel_abs
Label: Welfare aggregate used for national poverty estimation (annual)
Type: Numeric continuous variable
Description: This is the welfare aggregate used by the country to estimate its national poverty.
This aggregate can be nominal or spatially/temporally deflated. It should equal one of these four variables: pc_hh, padq_hh, pc_hhdr, padq_hhdr.

Use this code to generate wel_abs:

```
gen wel_abs = .
if wel_abs_deflation==0 & wel_abs_pcpadq==0 {
	replace wel_abs = pc_hh
}
if wel_abs_deflation==0 & wel_abs_pcpadq==1 {
	replace wel_abs = padq_hh
}
if inlist(wel_abs_deflation,1,2,3) & wel_abs_pcpadq==0 {
	replace wel_abs = pc_hhdr
}
if inlist(wel_abs_deflation,1,2,3) & wel_abs_pcpadq==1 {
```

```
	replace wel_abs = padq_hhdr
}
```

Variable: wel_fd

Label: Food part of welfare aggregate used for national poverty estimation (annual)
Type: Numeric continuous variable
Description:
This is the food part of the welfare aggregate used by the country to estimate its national poverty.
This aggregate can be nominal or spatially/temporally deflated. It should equal one of these four variables: pc_fd, padq_fd, pc_fddr, padq_fddr.
Use this code to generate wel_fd:

```
gen wel_fd = .
if wel_abs_deflation==0 & wel_abs_pcpadq==0 {
	replace wel_fd = pc_fd
}
if wel_abs_deflation==0 & wel_abs_pcpadq==1 {
	replace wel_fd = padq_fd
}
if inlist(wel_abs_deflation,1,2,3) & wel_abs_pcpadq==0 {
	replace wel_fd = pc_fddr
}
if inlist(wel_abs_deflation,1,2,3) & wel_abs_pcpadq==1 {
	replace wel_fd = padq_fddr
}
```

Variable: pl_abs

Label: National Absolute Poverty line (annual)
Type: Numeric continuous variable
Description: Country-derived by the NSO. If this variable is missing for some observations, replace missing values with the correct value.

```
levelsof(pl_abs)
if r(r)== 1 {
	if mi(pl_abs) replace pl_abs=`r(levels)'
}
else {
	display as error "pl_abs typically does not have multiple levels. Verify that this is
correct."
	exit
}
```

Variable: pl_fd

Label: National Food Poverty line (annual)
Type: Numeric continuous variable
Description: Country-derived by the NSO.

Variable: pl_ext

Label: National Hardcore poverty line (annual)
Type: Numeric continuous variable
Description: Country derived by the NSO. This line may be identical to the food poverty line or may be different.

Variable: poor_abs

Label: Absolute poor based on pl_abs
Type: Numeric categorical variable

Description:
Use this code to generate poor_abs: `gen poor_abs = wel_abs<pl_abs if !mi(wel_abs)`
1 = Poor
0 = Non-poor

Variable: poor_fd
Label: Food poor based on pl_fd
Type: Numeric categorical variable
Description:
Use this code to generate poor_fd: `gen poor_fd = wel_fd<pl_fd if !mi(wel_fd)`
1 = Poor
0 = Non-poor

Variable: poor_ext
Label: Hard core (extreme) poor based on pl_ext
Type: Numeric categorical variable
Description: Use this code to generate poor_ext: `poor_ext = wel_abs<pl_ext if !mi(wel_ext)`
1 = Poor
0 = Non-poor

Variable: converfactor
Label: Conversion factor
Type: Numeric continuous variable
Description: Specifies value for additional conversion factors if needed (e.g. from US$ to LCUs; currency change).

Variable: wel_PPPnom
Label: Welfare aggregate used for international poverty estimation (nominal, annual)
Type: Numeric continuous variable
Description: This is the nominal expenditure welfare aggregate.
This should equal pc_hh.
Use this code to generate wel_PPPnom: `gen wel_PPPnom = pc_hh`

Variable: wel_PPPdr
Label: Welfare aggregate used for international poverty estimation (deflated, annual)
Type: Numeric continuous variable
Description: This is the spatial and/or temporal deflated expenditure welfare aggregate.
This should equal pc_hhdr.
Use this code to generate wel_PPPdr: `gen wel_PPPdr = pc_hhdr`

Variable: wel_PPP_deflation
Label: Spatial/temporal deflation used for international poverty estimation
Type: Numeric categorical variable
Description:
0 = Neither spatially nor temporally deflated
1 = Spatially deflated
2 = Temporally deflated
3 = Both spatially and temporally deflated

Variable: wel_shpr
Label: Welfare aggregate for shared prosperity (if different from poverty)

Type: Numeric continuous variable
Description: This variable is for the welfare variable used to compute the shared prosperity indicator (e.g. per capita consumption) in the data file. This variable should be annual and in LCU at current prices. This variable is either the same as welfare (if same welfare aggregate is used for poverty and shared prosperity) or different if a different welfare aggregate is used for shared prosperity). In nearly all cases this variable will equal wel_PPP.

Variable: wel_shprtype
Label: Welfare type for shared prosperity indicator (income, consumption or expenditure)
Type: String variable
Description: Specifies the type of welfare measure for the variable welfshprosperity. Accepted values are: INC for income, CONS for consumption, or EXP for expenditure. Upper case must be used.

Variable: wel_oth
Label: Welfare aggregate if different welfare type is used from wel_abs, wel_PPPnom, wel_PPPdr
Type: Numeric continuous variable
Description: This variable is for the welfare aggregate in the data file if a different welfare type is used from the variables wel_abs, wel_PPPnom, wel_PPPdr. For example, if consumption is used for wel_abs, wel_PPPnom, wel_PPPdr but income also exists, it could be included here. This variable should be annual and in LCU at current prices.

Variable: wel_othtype
Label: Type of welfare measure (income, consumption or expenditure) for wel_oth
Type: String variable
Description: This variable specifies the type of welfare measure for the variable welfareother. Accepted values are: INC for income, CONS for consumption, or EXP for expenditure. This variable is only entered if the type of welfare is different from what is provided in wel_abs, wel_PPPnom, wel_PPPdr. For example, if consumption is used for wel_abs, wel_PPPnom, wel_PPPdr but income also exists, it could be included here. Welfaretype is case-sensitive and upper case must be used.

Variable: wel_PPP
Label: Welfare aggregate used for international poverty estimation (annual)
Type: Numeric continuous variable
Description: This is the final welfare variable used for international poverty monitoring purposes, that feeds into the GMD. It should equal either wel_PPPnom or wel_PPPdr.
Use this code to generate wel_PPP

```
gen wel_PPP = .
if wel_PPP_deflation==0 {
       replace wel_PPP = wel_PPPnom
}
if inlist(wel_PPP_deflation,1,2,3) {
       replace wel_PPP = wel_PPPdr
}
```

3 H Module – Household-level variables

The H-module contains household-level information (other than poverty) and includes information on housing characteristics and utilities, access to various amenities measured in terms of distances/time, and ownership of durable goods among others. To the extent possible, variables in this module should be generated independently from the I module. If necessary, you can copy code to generate the basic demographic variables.

3.1 Sample and Basic Household Identifiers

Variable: country
Label: Country code
Type: String variable
Description: 3-character length (Annex IV)

Variable: year_IHSN
Label: 4-digit year of survey based on IHSN standards
Type: Numeric discrete variable
Description: This is the start year of survey based on the IHSN standards. It should be identical to the year used for file-naming purposes.

Variable: hhno
Label: Household number
Type: Numeric discrete variable
Description: Household number

Variable: hid
Label: Household unique identification
Type: String or numeric variable
Description: This variable should uniquely identify observations and cannot be missing, i.e. isid hid should return no error.

Variable: wta_hh
Label: Household weights
Type: Numeric continuous variable
Description: To obtain household estimates, this is the weight to be used in all computations referring to household-level estimates. The interpretation is the proportion of households with a certain characteristic is XX%.

3.2 Housing and Utilities

Variable: ownhouse
Label: Ownership of dwelling unit
Type: Numeric categorical variable
Description: ownhouse is a categorical variable that specifies whether a household owns, rents, is provided for free, or squats in their house. Ownership (1) includes ownership or other equivalent of secure tenure, whether or not full-payment has been made yet. Rental (2) denotes that regular payment is made

to the owner (which could be private, corporate, or government) with or without formal agreement. This variable has four categories after harmonization:

1 = ownership/ secure rights
2 = renting
3 = provided for free
4= without permission

Variable: acqui_house

Label: Acquisition of house
Type: Numeric categorical variable
Description: acqui_house is a categorical variable that specifies the mode of acquisition for their dwellings. Only for household owners (Category 1 in ownhouse variable). Three categories after harmonization:
1 = Purchased; 2 = Inherited; 3 = Other
Category 3 would apply to cases if the members built their own homes or obtained it from other means specific to countries.

Variable: acqui_land

Label: Acquisition of land
Type: Numeric categorical variable
Description: acqui_land is a categorical variable that specifies the mode of acquisition for any residential land that the household uses. Only for the main residence. Only for land owners (category 1 in ownland variable). Three categories after harmonization:
1 = Purchased; 2 = Inherited; 3 = Other

Variable: dwelownlti

Label: Legal title for Ownership
Type: Numeric categorical variable
Description: dwelownlti is a dummy variable specifying whether a household has legal evidence for ownership (yes/no). Two categories after harmonization:
0 = No; 1 = Yes

Variable: dwelownti

Label: Type of Ownership Title
Type: Numeric categorical variable
Description: dwelownti is a categorical variable that specifies the type of legal document the household has as evidence for ownership of their dwelling. Type of legal document, six categories after harmonization:
1= Title, deed, freehold
2= Government issued leasehold
3= Occupancy certificate – govt issued
4= legal document in the name of group (community; cooperative)
5= condominium (apartment)
6= Other

Variable: fem_dwelownlti

Label: Legal title for Ownership
Type: Numeric categorical variable

Description: fem_dwelownlti is a dummy variable that specifies whether the names of female household members are listed on the legal document specifying ownership of the dwelling (yes/no). This will be derived from questions asking about the roster ID of the household member(s) whose name(s) are on the legal document for the dwelling. Two categories after harmonization:
0 = No; 1 = Yes

Variable: selldwel

Label: Right to sell dwelling
Type: Numeric categorical variable
Description: selldwel is a dummy variable that specifies whether the respondent has alienation rights (i.e. the right to sell) for their dwelling (yes/no). Two categories after harmonization:
0 = No; 1 = Yes

Variable: transdwel

Label: Right to transfer dwelling
Type: Numeric categorical variable
Description: transdwel is a dummy variable that specifies whether the respondent has the right to bequeath the dwelling to the next generation of their family (yes/no). Two categories after harmonization:
0 = No; 1 = Yes

Variable: ownland

Label: Ownership of land
Type: Numeric categorical variable
Description: ownland is a dummy variable that specifies whether a household owns residential land (yes/no). Ownership for property versus residential land on which property is constructed can be different in certain jurisdictions (land vested in a state or municipality). Two categories after harmonization:
0 = No; 1 = Yes

Variable: doculand

Label: Legal document for residential land
Type: Numeric categorical variable
Description: doculand is the dummy variable specifying whether the household has a legal document for their residential land (yes/no). Only for land owners (category 1 in ownland variable). Two categories after harmonization:
0 = No; 1 = Yes

Variable: fem_doculand

Label: Legal document for residential land – female
Type: Numeric categorical variable
Description: fem_doculand is the dummy variable specifying whether the household has the name of female household members listed on a legal document for their residential land (yes/no). This will be derived from questions asking about the roster ID of the household member(s) whose name(s) are on the legal document for residential land. Only for land owners (category 1 in ownland variable). Two categories after harmonization:
0 = No; 1 = Yes

Variable: landownti

Label: Land Ownership

Type: Numeric categorical variable

Description: landownti is a categorical variable that specifies the type of document that a household has to prove land ownership. The two customary rights categories (3 and 4) differentiate whether issued by plot or as a joined group title. Customary groups and cooperatives are differentiated, as well. Customary groups not required to have formal membership declared, while cooperative members have formalized status.Land ownership type of document. Only for land owners (category 1 in ownland variable). If the household owns multiple plots, this question should refer to the most common title type by area. Six categories after harmonization:

1 = Title; deed
2 = leasehold (govt issued)
3 = Customary land certificate/plot level
4 = Customary based / group right
5 = Cooperative group right
6 = Other

Use code that resembles the following:

```
collapse (sum) area, by(hhid category) //keeps only 1 obs per hhid/category/plot
collapse (max) area, by(hhid category) //keeps only 1 obs per hhid/category
bysort hhid: egen _temp=max(area) //creates a temporary variable _temp with max area
keep if _temp==area & category=.  //keeps only 1 obs per hhid
```

Variable: sellland

Label: Right to sell land

Type: Numeric categorical variable

Description: sellland is a dummy variable that specifies whether the respondent has alienation rights (i.e. the right to sell) for their residential land (yes/no). Only for land owners (category 1 in ownland variable). Two categories after harmonization:

0 = No; 1 = Yes

Variable: transland

Label: Right to transfer land

Type: Numeric categorical variable

Description: transland is a dummy variable that specifies whether the respondent has the right to bequeath residential land to the next generation of their family (yes/no). Only for land owners (category 1 in ownland variable). Two categories after harmonization:

0 = No; 1 = Yes

Variable: agriland

Label: Agriculture Land

Type: Numeric categorical variable

Description: agriland is a dummy variable that specifies whether a household is using agricultural land according to the classification of the World Census of Agriculture 2020.[1] Two categories after harmonization: 0 = No; 1 = Yes

[1] FAO (2015). "WORLD PROGRAMME FOR THE CENSUS OF AGRICULTURE 2020". Paragraph (8.2.35) FAO's recommended land use classification in the Figure 1 includes the following aggregate classes:

- **Arable land** is land that is used in most years for growing temporary crops. It includes land used for growing temporary crops during a twelve-month reference period, as well as land that would normally be

Variable: area_agriland

Label: Area of agriculture land used (in hectares)

Type: Numeric continuous variable

Description: area_ownagriland is a numeric, continuous variable that specifies the total area of agricultural land used in hectares. This could be land that is owned, rented, or sharecropped, or some combination. A hectare is equal to 10,000 square meters or equivalent to 2.471 acres.

Variable: ownagriland

Label: Ownership of agriculture land

Type: Numeric categorical variable

Description: ownagriland is a dummy variable that specifies whether a household owns agricultural land (yes/no). Owned land can be by freehold, deed, customary, or government leasehold. Only those households that declared using agricultural land (category 1 in agriland variable). Two categories after harmonization:

0 = No; 1 = Yes

Variable: area_ownagriland

Label: Area of agriculture land owned (in hectares)

Type: Numeric continuous variable

Description: area_ownagriland is a numeric, continuous variable that specifies the total area of agricultural land owned in hectares. Only for agriculture land owners (category 1 in ownagriland variable). A hectare is equal to 10,000 square meters or equivalent to 2.471 acres.

Variable: purch_agriland

Label: Purchased agri land

Type: Numeric categorical variable

Description: purch_agriland is a dummy variable specifying whether a household has purchased the agricultural land they own (yes/no). Only for agriculture land owners (category 1 in ownagriland variable). Two categories after harmonization:

0 = No; 1 = Yes

Variable: areapurch_agriland

Label: Area of purchased agriculture land (in hectares)

---

so used but is lying fallow or has not been sown due to unforeseen circumstances. Arable land does not include land under permanent crops or land that is potentially cultivable but is not normally cultivated. Such land should be classified as “permanent meadows and pastures” if used for grazing or haying, “forest and other wooded land” if overgrown with trees and not used for grazing or haying, or “other area not elsewhere classified” if it becomes wasteland.

- **Cropland** is the total of arable land and land under permanent crops.
- **Agricultural land** is the total of cropland and permanent meadows and pastures.
- **Land used for agriculture** is the total of “agricultural land” and “land under farm buildings and farmyards”.

0203 Area of holding according to land tenure types

- Legal ownership or legal owner-like possession
- Non-legal ownership or non-legal owner-like possession
- Rented from someone else

Other types of land tenure

Type: Numeric continuous variable
Description: areapurch_agriland is a numeric, continuous variable that specifies the total area of agricultural land purchased in hectares. Only for category 1 in purch_agriland variable. A hectare is equal to 10,000 square meters or equivalent to 2.471 acres.

Variable: inher_agriland
Label: Inherit agriculture land
Type: Numeric categorical variable
Description: inher_agriland is a dummy variable specifying whether a household has inherited the agricultural land they own (yes/no). Only for agriculture land owners (category 1 in ownagriland variable). Two categories after harmonization:
0 = No; 1 = Yes

Variable: areainher_agriland
Label: Area of inherited agriculture land (in hectares)
Type: numeric continuous variable
Description: areainher_agriland is a numeric, continuous variable that specifies the total area of agricultural land inherited in hectares. Only for category 1 in inher_agriland variable. A hectare is equal to 10,000 square meters or equivalent to 2.471 acres.

Variable: rentout_agriland
Label: Rent Out Land
Type: Numeric categorical variable
Description: rentout_agriland is a dummy variable that specifies whether any of the agricultural land a household uses is rented–out land or sharecropped (yes/no). Only for agriculture land owners (category 1 in ownagriland variable). This refers to land (or use rights) owned by the household but cultivated or utilized by someone else irrespective of the type of the tenant (individual, household, legal entity, etc.) and contractual arrangements (fixed rental, sharecropping, etc.). Two categories after harmonization:
0 = No; 1 = Yes

Variable: arearentout_agriland
Label: Area of rent out agri land (in hectares)
Type: Numeric continuous variable
Description: arearentout_agriland is a numeric, continuous variable that specifies the total area of agricultural land rented out or share cropped in hectares. Only for category 1 in rentout_agriland variable. A hectare is equal to 10,000 square meters or equivalent to 2.471 acres.

Variable: rentin_agriland
Label: Rent in Land
Type: Numeric categorical variable
Description: rentin_agriland is a dummy variable that specifies whether any of the agricultural land a household uses is rented–in land or sharecropped (yes/no). This refers land owned by others (not members of the household) but cultivated or used by the household under fixed rental, sharecropped or similar arrangements. We agree that this question should apply to all households using agricultural land (agriland==1). Two categories after harmonization:
0 = No; 1 = Yes

Variable: arearentin_agriland

Label: Area of rent in agri land (in hectares)
Type: Numeric continuous variable
Description: arearentin_agriland is a numeric, continuous variable that specifies the total area of agricultural land rented in or share cropped in hectares. Only for category 1 in rentin_agriland variable. A hectare is equal to 10,000 square meters or equivalent to 2.471 acres

Variable: docuagriland
Label: Documented Agri Land
Type: Numeric categorical variable
Description: docuagriland is the dummy variable specifying whether the household has a legal document for their agricultural land (yes/no). Only for agriculture land owners (category 1 in ownagriland variable). Two categories after harmonization:
0 = No; 1 = Yes

Variable: area_docuagriland
Label: Area of documented agri land (in hectares)
Type: Numeric continuous variable
Description: Area_docuagriland is a numeric, continuous variable that specifies the total area of agricultural land owned with legal documentation in hectares. Only for category 1 in docuagriland variable. A hectare is equal to 10,000 square meters or equivalent to 2.471 acres.

Variable: fem_agrilandownti
Label: Ownership Agri Land – Female
Type: Numeric categorical variable
Description: fem_agrilandownti is the dummy variable specifying whether the household has the name of female household members listed on a legal document for their agricultural land (yes/no). This will be derived from questions asking about the roster ID of the household member(s) whose name(s) are on the legal document for agricultural land. Only for category 1 in docuagriland variable. Two categories after harmonization:
0 = No; 1 = Yes

Variable: agrilandownti
Label: Type Agri Land ownership doc
Type: Numeric categorical variable
Description: agrilandownti is a categorical variable that specifies the type of document that a household has to prove agricultural land ownership. The two customary rights categories (3 and 4) differentiate whether issued by plot or as a joined group title. Customary groups and cooperatives are differentiated, as well. Customary groups not required to have formal membership declared, while cooperative members have formalized status. Agricultural land ownership type of document. Only for category 1 in docuagriland variable. If the household owns multiple plots, this question should refer to the most common title type by area. Categories after harmonization:
1 = Title; deed
2 = leasehold (govt issued)
3 = Customary land certificate/plot level
4 = Customary based / group right

5 = Cooperative
6 = Other

Variable: sellagriland

Label: Right to sell agri land

Type: Numeric categorical variable

Description: sellagriland is a dummy variable that specifies whether the respondent has alienation rights (i.e. the right to sell) for their agricultural land (yes/no). Only for agricultural land owners, category 1 in ownagriland variable. Two categories after harmonization: 0 = No; 1 = Yes

Variable: transagriland

Label: Right to transfer agri land

Type: Numeric categorical variable

Description: transagriland is a dummy variable that specifies whether the respondent has the right to bequeath agricultural land to the next generation of their family (yes/no). Only for agricultural land owners, category 1 in ownagriland variable. Two categories after harmonization:
0 = No; 1 = Yes

Variable: typlivqrt

Label: Types of living quarters

Type: Numeric categorical variable

Description: typlivqrt is a categorical variable that specifies the type of living quarters. Categories after harmonization are:
1 = Housing units, conventional dwelling with basic facilities
2 = Housing units, conventional dwelling without basic facilities
3 = Other

Variable: dweltyp

Label: Types of Dwelling

Type: Numeric categorical variable

Description: dweltyp is a categorical variable that specifies the type of dwelling. Categories after harmonization are:

1 = Detached house;
2 = Multi-family house
3 = Separate apartment;
4 = Communal apartment
5 = Room in a larger dwelling;
6 = Several buildings connected
7 = Several separate buildings;
8 = Improvised housing unit
9 = Other

Variable: ybuilt

Label: Year the dwelling built

Type: Numeric discrete variable

Description: ybuilt is an integer variable that indicates the year when the dwelling was built, regardless of the ownership status.

Variable: rooms

Label: Number of habitable rooms

Type: Numeric discrete variable

Description: rooms is an integer variable that refers to the number of habitable rooms in the whole household dwelling unit. It may consist of one or more structure(s) (rooms), including all rooms used for living, sleeping and eating. It excludes storerooms, bathrooms, kitchens and rooms used for business or professional purposes. In the case of a one-room dwelling this variable will have the value of one.

Variable: areaspace

Label: Area

Type: Numeric continuous variable

Description: areaspace is a continues variable that refers to the total floor area (in square meters) of all rooms and auxiliary premises (kitchen, vestibule, cloakroom, hallway, toilet room, sauna that is within the dwelling, pantry, interstice, bathroom, storeroom, porch, integrated wall closets) in the whole household dwelling unit. The area of the dwelling does not include cellars, garages (incl. in private houses), boiler rooms, attics (if they are not suitable for permanent habitation) and common rooms (such as stairways, corridors, saunas, etc.) in buildings with multiple dwellings. Open areas (loggias, balconies and terraces) are not included in the area of the dwelling. However, if such areas have been closed in and insulated, they should be added to the total area of the dwelling. If a household lives in an uncompleted residential building, enter the area of the finished part of the house.

Variable: roofcs

Label: Main material used for roof (country specific)

Type: String variable

Description: This refers to the variable on roof material (if any), as it comes in the survey. If more than one material is used for structure, the dominant material is the information required. The format should be code and value label. For example, "1 - Stone"; "2 - Mud"; etc.

Variable: roof

Label: Main material used for roof

Type: Numeric categorical variable

Description: roof is a categorical variable that indicates type of material used for roof, such as adobe, thatch, iron, and tiles. The roof material is categorized into 3 broad categories namely: Natural, rudimentary and finished. For cases that cannot be covered in the above three categories, please use code 15 = Other – "Specific".

1 = Natural – Thatch/palm leaf;
2 = Natural – Sod;
3 = Natural – Other;
4 = Rudimentary – Rustic mat;
5 = Rudimentary – Palm/bamboo;
6 = Rudimentary – Wood planks;
7 = Rudimentary – Other;
8 = Finished – Wood;
9 = Finished – Asbestos;
10 = Finished – Tile;
11 = Finished – Concrete;
12 = Finished – Metal;
13 = Finished – Roofing shingles;
14 = Finished – Other
15 = Other

Variable: wallcs

Label: Main material used for external walls (country specific)

Type: String variable

Description: This refers to the variable on external wall material (if any), as it comes in the survey. If more than one material is used for structure, the dominant material is the information required. The format should be code and value label. For example, "1 - Stone"; "2 - Mud"; etc

Variable: wall

Label: Main material used for external walls

Type: Numeric categorical variable

Description: wall is a categorical variable that indicates type of material used for walls. The wall material is categorized into 3 broad categories namely: Natural, rudimentary and finished. For cases that cannot be covered in the above three categories, please use code 19 = Other – "Specific". Main source of material used for walls, 19 categories after harmonization:

1 = Natural – Cane/palm/trunks; 2 = Natural – Dirt
3 = Natural – Other; 4 = Rudimentary – Bamboo with mud
5 = Rudimentary – Stone with mud; 6 = Rudimentary – Uncovered adobe
7 = Rudimentary – Plywood; 8 = Rudimentary – Cardboard
9 = Rudimentary – Reused wood; 10 = Rudimentary – Other
11 = Finished – Woven Bamboo; 12 = Finished – Stone with lime/cement
13 = Finished – Cement blocks; 14 = Finished – Covered adobe
15 = Finished – Wood planks/shingles; 16 = Finished – Plaster wire
17 = Finished – GRC/Gypsum/Asbestos; 18 = Finished – Other
19 = Other

Variable: floorcs

Label: Main material used for floor (country specific)

Type: String variable

Description: This refers to the variable on floor material (if any), as it comes in the survey. If more than one material is used for structure, the dominant material is the information required. Format should be code and value label. For example, "1 - Stone"; "2 - Mud"; etc

Variable: floor

Label: Main material used for floor

Type: Numeric categorical variable

Description: floor is a categorical variable that indicates type of material used for floors. The floor material is categorized into 3 broad categories namely: Natural, rudimentary and finished. For cases that cannot be covered in the above three categories, please use code 14 = Other – "Specific".

Main source of material used for floors, 14 categories after harmonization as shown below.

1 = Natural – Earth/sand; 2 = Natural – Dung;
3 = Natural –¬ Other; 4 = Rudimentary –¬ Wood planks
5 = Rudimentary –¬ Palm/bamboo; 6 = Rudimentary – Other
7 = Finished – Parquet or polished wood; 8 = Finished – Vinyl or asphalt strips
9 = Finished – Ceramic/marble/granite; 10 = Finished – Floor tiles/terrazzo
11 = Finished – Cement/red bricks; 12 = Finished – Carpet
13 = Finished – Other; 14 = Other

Variable: watercs_type

Label: Type of water questions used in the survey

Type: Numeric categorical variable

Description: This variable records the type of question(s) asked about access to water in the survey. For example, if the survey had a specific question on the water source on drinking water, or on water source on general water, or both. Subsequent question on water will depend on this response.

Four categories after harmonization:

1 = Drinking water; 2 = General water; 3 = Both; 4 = Other

Variable: watercs

Label: Main source of water (country specific)

Type: String variable

Description: This refers to the variable on the main water source (if any), as it comes in the survey. If more than one water source, only main source required. In some surveys, drinking water is asked and is differentiated from other water uses. In these cases, use the drinking water source to code this variable. If two sources of water are available (water source during the wet and dry season), use water source during dry season. The reason for using water during the dry season is that the world is experiencing global warming and the climate is changing rapidly. The format should be code and value label. For example, "1 - Pipe"; "2 - Spring"; etc.

Variable: watercs_d

Label: Main source of water during the dry season (country specific)

Type: String variable

Description: Question must be explicitly asked in survey on water source during the dry season.
Labels must be translated to English. Make sure translation is correct from a language expert.
If more than one water source, only main source required.
In some surveys, drinking water is asked and is differentiated from other water uses. Use the drinking water source to code this variable. For each value label, there should be a space between the hyphen.
Format should be code and value label. For example, "1 – Pipe"; "2 – Spring"; etc.

Variable: water14

Label: Main source of drinking water (14 categories)

Type: Numeric categorical variable

Description: Water14 is a categorical variable that indicates the main source of drinking water for the household. If the main source of water differs between the wet and dry season, water source during dry season is referred. The best possible match is sought, but in many cases the correspondence between country-specific values and these standardized codes is imperfect. You should refer to the survey questionnaire to assess the best matches. Category 7 (bottled water) includes all forms of packaged water including bottles and sachets.

1 = Piped water into dwelling;	2 = Piped water to yard/plot;
3 = Public tap or standpipe;	4 = Tubewell or borehole;
5 = Protected dug well;	6 = Protected spring;
7 = Bottled water;	8 = Rainwater;
9 = Unprotected spring;	10 = Unprotected dug well;
11 = Cart with small tank/drum;	12 = Tanker-truck;
13 = Surface water;	14 = Other

Variable: water8

Label: Main source of drinking water (8 categories)

Type: Numeric categorical variable

Description: Wells include springs, boreholes but must be protected from any possible sources of contamination such as surface water or seepage.

1 = Piped water (own tap);	2 = Public tap or standpipe
3 = Protected well;	4 = Unprotected well
5 = Surface water;	6 = Rainwater
7 = Tanker-truck, vendor;	8 = Other

```
recode water14 (1=1) (2 3=2) (4 5 6=3) (9 10=4) (13=5) (8=6) (11 12=7)
14=8),gen(water8); ta water14 water8
```

Variable: waterpipe

Label: Household has piped water

Type: Numeric categorical variable

Description: Main water source is piped water which can be within household, plot or public standpipe. "Piped" is the condition. Four categories after harmonization:

0 = No

1 = Yes, in premise

2 = Yes, but not in premise

3 = Yes, unstated whether in or outside premise

```
recode water14 (1 2=1) (3=2) (else=0), gen(waterpipe)
replace waterpipe=. if water14==.
```

If water14 is missing but you have the information to code waterpipe in watercs, do not use the code above. water14 does not have enough information to code category 3, thus you may need to use information from watercs to add this category.

Variable: piped

Label: Access to piped water

Type: Numeric categorical variable

Description: piped is a categorical variable that indicates whether the household has access to piped water. There are two major types of water supply – within premises and outside premises. 'Within premises' refers to water service piped connection to own tap. It includes both household connection (in-house plumbing) and yard connection (yard or plot outside the house plumbing). Conversely, outside premise refers to a public water point from which people can collect water, shared among houses. It includes public tap and standpipe or a public fountain. Two categories after harmonization:

0 = No; 1 = Yes (Piped water into dwelling, piped water to yard/plot, or public tap or standpipe)

```
gen piped = .
replace piped = 1 if inlist(water14,1,2,3)
replace piped = 0 if !inlist(water14,1,2,3,.)
```

Variable: piped_to_prem

Label: Access to piped water on premises

Type: Numeric categorical variable

Description: piped_to_prem is a categorical variable that specifies whether a household has access to piped water on premises. There are two major types of water supply – within premises and outside premises. 'Within premises' refers to water service piped connection to own tap. It includes both household connection (in-house plumbing) and yard connection (yard or plot outside the house plumbing). Conversely, outside premise refers to a public water point from which people can collect water, shared among houses. It includes public tap and standpipe or a public fountain.

```
gen piped_to_prem = .
replace piped_to_prem = 1 if inlist(water14,1,2)
replace piped_to_prem = 0 if !inlist(water14,1,2,.)
```

Variable: imp_wat_rec

Label: Household has improved water sources

Type: Numeric categorical variable

Description: When possible, this variable should be derived from the variable water14, with categories 1-6 and 8 as Yes (1) and other categories as No (0), the last option (14) can be very country-specific judgement to the definition of improved access to water. Bottled water is an improved source only if accompanied by another improved source. When there is no water source variable or the categorical responses from the survey cannot be mapped into the water sources, you might still be able to map into improved access to water based on country specific information. Often, the JMP data excel file is a good source of cross-validation on this variable harmonization (https://washdata.org/data#!/). Another useful source (https://www.cdc.gov/healthywater/global/assessing.html) for assessing whether a category is improved. Use the following code:

```
recode water14 (1/6 8=1) (nonmissing=0), gen(imp_wat_rec)
```

imp_wat_rec is a categorical variable that estimates the "recommended" categorization for access to improved water sources in each country, or how evidence suggests that the expected error might be minimized. If the relevant survey was on file in the SDG calculations, this would be considered 1 if the majority of the problematic category was estimated therein to be of an improved type at the rural level, and otherwise considered 0. If the survey was not already in the SDG calculations, recommendations are based on the standard international classifications plus any relevant insights from other surveys on file for the specific country. In the few instances where there was no evidence, 0 is used. *To harmonize this variable, use the classification from the WASH Team.* Two categories after harmonization:
0 = No; 1 = Yes

Variable: w_30m

Label: Access to water within 30 minutes
Type: Numeric categorical variable
Description: w_30m is a categorical variable that specifies whether a household has access to improved water within 30 minutes. This includes time taken for a round trip and waiting time in case of queues. This variable needs to be created in conjunction with the imp_wat_rec dummy to identify where the improved water source is available within 30 minutes. Collection time of imp_wat_rec within 30 minutes, two categories after harmonization:
1=collection time of imp_wat_rec less than or equal to 30 mins;
0=collection time of imp_wat_rec more than 30 mins

Variable: w_avail

Label: Water is available when needed
Type: Numeric categorical variable
Description: w_avail is a categorical variable that specifies whether improved water is available when needed. This variable needs to be created in conjunction with the imp_wat_rec dummy to identify where the improved water source is available reliably 24/7. Categories after harmonization:
1= water is available continuously, reliable source
0=water source is unreliable

Variable: adiswat_d

Label: Actual distance to main water point (kms) during the dry season
Type: Numeric continuous variable
Description: This refers to actual distance to water point (one way) used by household in kms during the dry season. If no season is specified, use this variable.
By convention: 1 km = 1000 m; 1 km = 5/8 mile. If within dwelling, code zero.

Variable: adiswat_w

Label: Actual distance to main water point (kms) during the wet season
Type: Numeric continuous variable
Description: This refers to actual distance to water point (one way) used by household in kms.
By convention: 1 km = 1000 m; 1 km = 5/8 mile.
If within dwelling, code zero.

Variable: atimwat_d

Label: Actual time taken to main water point (mins) during the dry season
Type: Numeric continuous variable
Description: This refers to actual time taken to water point used by household.
If roundtrip provided, divide by 2.

Variable: atimwat_w

Label: Actual time taken to main water point (mins) during the wet season
Type: Numeric continuous variable
Description: This refers to actual time taken to water point used by household.
If roundtrip provided, divide by 2.

Variable: toiletcs

Label: Main toilet facility (country specific)
Type: string variable
Description: Labels must be translated to English. Make sure translation is correct from a language expert.
For each value label, there should be a space between the hyphen.
Format should be code and value label. For example, "1 – Flush"; "2 – VIP"; etc.

Variable : toilet14

Label : Main toilet facility (14 categories)
Type: Numeric categorical variable
Description: sanitation_source is a categorical variable that specifies the source of sanitation facilities. The best possible match is sought, but in many cases the correspondence between country-specific values and these standardized codes is imperfect. You should refer to the survey questionnaire to assess the best matches.
Main sanitation source, fourteen categories after harmonization:

1 = A flush toilet;	2 = A piped sewer system
3 = A septic tank;	4 = Pit latrine
5 = Ventilated improved pit latrine (VIP);	6 = Pit latrine with slab
7 = Composting toilet;	8 = Special case
9 = A flush/pour flush to elsewhere;	10 = A pit latrine without slab
11 = Bucket;	12 = Hanging toilet or hanging latrine
13 = No facilities or bush or field;	14 = Other

Category 8 applies to improved sanitation facilities for which the respondent does not know whether the facility is connected to a sewer or septic tank.

Variable: toilet6

Label: Main toilet facility (6 categories)
Type: Numeric categorical variable
Description: Must be coded from toilet14.
1 = Flush toilet;

2 = Ventilated Improved Pit (VIP) latrine
3 = Composting toilet;
4 = Pit latrine with slab
5 = No facility;
9 = Other
The code for generating toilet6:

```
recode toilet14 (1/3=1) (5=2) (7=3) (6=4) (13=5) (else=9),gen(toilet6)
replace toilet6=. if toilet14==.
```

Variable: toiletflush

Label: Access to flushed toilet
Type: Numeric categorical variable
Description: Must be asked in survey explicitly. Do not guestimate.
0 = No
1 = Yes, in premise
2 = Yes, but not in premise including public toilet
3 = Yes, unstated whether in or outside premise

Variable: sewer

Label: sewer
Type: Numeric categorical variable
Description: sewer is a categorical variable that specifies whether a household has access to a toilet connected to a piped sewer system. Access to sewer, two categories after harmonization:
0 = No
1 = flush/pour flush to piped sewer system

Variable: open_def

Label: Access to any sanitation facility
Type: Numeric categorical variable
Description: open_def is a categorical variable that specifies whether a household has access to any sanitation facility. Two categories after harmonization:
0=availability of any facility (from list of categories in sanitation_source including unimproved options)
1=no facility, or bush, or field (13)
Code to create this variable when toilet14 is available in the dataset:

```
recode toilet14 (13 14=1) (else=0), gen(open_def)
replace open_def=. if toilet14==.
```

Variable: toiletshared

Label: toilet facility shared with other households
Type: Numeric categorical variable
Description: This question must have been asked in the survey.
If question not asked leave as missing.
0 = No; 1 = Yes

Variable: imp_san_rec

Label: access to improved sanitation
Type: Numeric categorical variable

Description: This includes toilet6<=4 and not shared. imp_san_rec is a categorical variable that estimates the categorization for access to improved sanitation facilities in each country, or how evidence suggests that the expected error might be minimized. If the relevant survey was on file in the SDG calculations, this would be considered 1 if the majority of the problematic category was estimated therein to be of an improved type at the rural level, and otherwise considered 0. If the survey was not already in the SDG calculations, recommendations are based on the standard international classifications plus any relevant insights from other surveys on file for the specific country. In the few instances where there was no evidence, 0 is used. If question of shared toilet facility is asked, use the variable to recode appropriately. *To harmonize this variable, use the classification from the WASH Team.* Another useful source for assessing whether a category is improved (https://www.cdc.gov/healthywater/global/assessing.html). Use the following Stata code:

```
recode toilet6 (1/4=1) (nonmissing=0), gen(imp_san_rec)
replace imp_san_rec=0 if toiletshared==1
```

Two categories after harmonization:
0 = No; 1 = Yes

Variable: fuelcookcs

Label: Main cooking fuel (country specific)
Type: String variable
Description: If several fuels asked in survey, only main source required.
Labels must be translated to English. Make sure translation is correct from a language expert.
For each value label, there should be a space between the hyphen.
Format should be code and value label. For example, "1 – Electricity"; "2 – Firewood"; etc.

Variable: fuelcook

Label: Main cooking fuel
Type: Numeric categorical variable
Description: fuelcook a categorical variable that identifies the source of cooking.
1 = Firewood
2 = Kerosene
3 = Charcoal
4 = Electricity
5 = Gas
9 = Other
10 = None

Variable: fuellighcs

Label: Main lighting fuel (country specific)
Type: String variable
If several fuels asked in survey, only main source required.
Labels must be translated to English. Make sure translation is correct from a language expert.
For each value label, there should be a space between the hyphen.
Format should be code and value label. For example, "1 – Electricity"; "2 – Firewood"; etc.

Variable: fuelligh

Label: Main lighting fuel

Type: Numeric categorical variable
Description: fuelligh is a categorical variable that identifies the source of light. The categories after harmonization are:
1 = Electricity
2 = Kerosene
3 = Candles
4 = Gas
9 = Other
10 = None

Variable: electyp

Label: Source of energy
Type: Numeric categorical variable
Description: electyp is a categorical variable that specifies the source of energy when fuelcook and fuelligh variables are not available and there is only one question about the type of energy source in the household; when fuelcook and fuelligh are available this variable has to be created prioritizing electricity, then Gas, then Lamp. Four categories after harmonization:
1 = Electricity
2 = Gas
3 = Lamp
4 = Others
10 = None

When fuelcook and fuelligh are available, electyp can be created using the following code:

```
gen electyp=.
replace electyp=1 if fuelcook==4 | fuelligh==1
replace electyp=2 if (fuelcook==5 | fuelligh==4) & mi(electyp)
replace electyp=3 if (fuelcook==2 | inlist(fuelligh,2,3)) & mi(electyp)
replace electyp=4 if (inlist(fuelcook,1,3,9) | fuelligh==9) & mi(electyp)
replace electyp=10 if fuelcook==10 & fuelligh==10
```

Variable: elecsource

Label: Main source of electricity
Type: Numeric categorical variable
Description: Use both FUELCOOK and FUELLIGH. FUELLIGH should be the main one to use.
If electricity source not specified, code “other” but this should be on a country-to-country situation.
1 = Mains; 2 = Solar; 3 = Generator; 4 = Other; 5 = No electricity

Variable: electricity

Label: Household has access to electricity
Type: Numeric categorical variable
Description: electricity is a dummy variable that specifies whether the household has access to electricity in the dwelling, irrespective of the source. Possible sources could be mains, solar, generator, etc. Categories after harmonization:
0 = No; 1 = Yes

Variable: elec_acc

Label: Access to electricity

Type: Numeric categorical variable
Description: elec_acc is a categorical variable that identifies type of connection to electricity. For instance, access to electricity ('Yes') may be public/quasi-public referring to mains electricity (i.e. the term used to refer to the electricity supply from power stations to households) or private referring to electricity from generator or solar or private company. The quality of electricity is assessed by other Tier 3 variables, such as number of electricity hours per day (elechr_acc). Categories after harmonization:
1 = Yes, public/quasi-public
2 = Yes, private
3 = Yes, source unstated
4 = No

Variable: elechr_acc
Label: Electricity availability (hr/day)
Type: Numeric continuous variable
Description: elechr_acc is a numeric continuous variable that specifies the access to electricity in hours per day.

Variable: kitchen
Label: Separate kitchen in dwelling
Type: Numeric categorical variable
Description: kitchen is a dummy variable indicating whether the household has a separate kitchen in the dwelling, implying an independent space is set aside for cooking inside the dwelling (kitchen). Any other space reserved for cooking, such as kitchenette or an outer space for kitchen, is not considered as a kitchen. The unit of enumeration for this topic is the housing unit. However, some countries may find it useful to collect information on the availability of kitchen facilities for the use of occupants in collective living quarters, such as hotels, lodging houses, institutions camps and workers' quarters, though people living in these places are generally not captured in a household survey. Two categories after harmonization:
0 = No; 1 = Yes

Variable: bath
Label: Bathing facility such as shower or bathtub in the dwelling
Type: Numeric categorical variable
Description: bath is a dummy variable indicating whether the household has a separate bathing facility such a shower or bathroom in the dwelling. Fixed bath or shower outside housing unit is not considered. Two categories after harmonization:
0 = No; 1 = Yes

Variable: garbdispcs
Label: Garbage and trash disposal (country specific)
Type: String variable
Description: Labels must be translated to English. Make sure translation is correct from a language expert. For each value label, there should be a space between the hyphen.
Format should be code and value label. For example, "1 – Collected"; "2 – Buried"; "3 - Street"; etc.

Variable: garbdisp

Label: Garbage and trash disposal
Type: Numeric categorical variable
Description: Refers to only garbage or trash generated by household.
1 = Collected
2 = Buried/burned
3 = Discarded in empty lots, street, rivers
9 = Other

Variable: garbdisp10

Label: Garbage and trash disposal
Type: Numeric categorical variable
Description: waste is a categorical variable that indicates the type of solid waste disposal. This variable contains information on the usual manner of collection and disposal of solid waste or garbage generated by occupants of the housing unit. Type of solid waste disposal is categorized by the manner of disposal, such as collection, disposal, burial or compost and by the administrator of the waste disposal, such as authorized collectors, self-appointed collectors, and dump supervised by authorities.

Main types of sewage disposal system, ten categories after harmonization:
1 = Solid waste collected on a regular basis by authorized collectors;
2 = Solid waste collected on an irregular basis by authorized collectors;
3 = Solid waste collected by self-appointed collectors;
4 = Occupants dispose of solid waste in a local dump supervised by authorities;
5 = Occupants dispose of solid waste in a local dump not supervised by authorities;
6 = Occupants burn solid waste;
7 = Occupants bury solid waste;
8 = Occupants dispose solid waste into river, sea, creek, pond;
9 = Occupants compost solid waste;
10 = Other arrangement.

Variable: central_acc

Label: Access to central heating
Type: Numeric categorical variable
Description: central_acc is a dummy variable that indicates the access to central heating in the dwelling.
Categories after harmonization:
0 = No; 1 = Yes

Variable: heatsource

Label: Main source of heating
Type: Numeric categorical variable
Description: heatsource is a categorical variable that indicates the main source of heating. Main source of heating refers to the type of system used to provide heating for most of the space. It may be central heating covering all or parts of living quarters, or it may not be central, in which case the heating will be provided separately within the living quarters by a stove, fireplace or some other heating body.

As for the energy used for heating purposes, it is closely related to the type of heating and refers to the predominant source of energy, such as solid fuels (coal, lignite, and products of coal and lignite, wood), oils, gaseous fuels (natural or liquefied gas), or electricity.
Main sources of heating, seven categories after harmonization:

1 = Firewood;	2 = Kerosene;
3 = Charcoal;	4 = Electricity;
5 = Gas;	6 = Central;
9 = Other	10 = No heating

Variable: gas

Label: Connection to gas/Usage of gas

Type: Categorical variable

Description: gas is a categorical variable that identifies type of gas usage. The categories after harmonization are:

0 = No
1 = Yes, piped gas (LNG)
2 = Yes, bottled gas (LPG)
3 = Yes, but don't know

3.3 Utilities Expenditures

The variables in this section should be expressed in current prices in the local currency unit (LCU) without any spatial or temporal deflation.

The table below summarizes all the utilities expenditure variables. The variables highlighted in yellow are secondary variables that are aggregated using primary variables. However, there might be surveys that report expenditures on secondary level only. For example: waste expenditure (waste_exp) is sum of garbage expenditure (garbage_exp) and sewage expenditure (sewage_exp). In surveys where expenditures are reported on disaggregated level will include values for garbage expenditure and sewage expenditure and then waste_exp is created by adding garbage and sewage expenditures. However, some surveys will report expenditure only for total waste i.e. waste_exp, leading to missing values for garbage_exp and sewage_exp.

Variable: pwater_exp

Label: Total annual consumption of water supply/piped water

Type: Numeric continuous variable

Description: pwater_exp is a continuous variable that refers to total annual household expenditures on water supply/piped water. It includes associated expenditure such as hire of meters, reading of meters, standing charges, etc. GMD water consumption variables include an aggregate water variable comprising water supply (pwater_exp) and hot water (hwater_exp) and defined as water_exp. As in the case of the COICOP classification, the variable excludes household expenditures on hot water. Drinking water sold in bottles or containers is also excluded from water supply.

Variable: hwater_exp

Label: Total annual consumption of hot water

Type: Numeric continuous variable

Description: hwater_exp is a continuous variable that refers to total annual household expenditure on hot water supply.

Variable: water_exp

Label: Total annual consumption of water supply and hot water

Type: Numeric continuous variable

Description: water_exp is a continuous variable that refers to total annual household expenditure on water supply and hot water supply. This variable specifies the sum of expenditure of water supply (pwater_exp) and hot water supply (hwater_exp).

Variable: garbage_exp

Label: Total annual consumption of garbage collection

Type: Numeric continuous variable

Description: garbage_exp is a continuous variable that refers to total annual household expenditures on collection and disposal of garbage or refuse.

Variable: sewage_exp

Label: Total annual consumption of sewage collection

Type: Numeric continuous variable

Description: sewage_exp is a continuous variable that refers to total annual household expenditures on collection and disposal of wastewater.

Variable: waste_exp

Label: Total annual consumption of garbage and sewage collection

Type: Numeric continuous variable

Description: waste_exp is a continuous variable that refers to the total annual household expenditure on garbage (garbage_exp) and sewage (sewage_exp) collection.

Variable: dwelothsvc_exp

Label: Total annual consumption of other services relating to the dwelling

Type: Numeric continuous variable

Description: dwelothsvc_exp is a continuous variable that refers to total annual household expenditures on other services relating to the dwelling. These expenditures typically include co-proprietor charges in multi-occupied buildings, security services, and other miscellaneous services. Co-proprietor charges include charges for caretaking, gardening, stairwell cleaning, heating and lighting, maintenance of lifts and refuse disposal chutes, etc. This variable does not include household services such as window cleaning, disinfecting, fumigation and pest extermination ; bodyguards . Maintenance and repair of the dwelling is also excluded from other services relating to the dwelling (dwelothsvc_exp) but included as additional variables defined as dwelmat_exp and dwelsvc_exp.

Variable: elec_exp

Label: Total annual consumption of electricity

Type: Numeric continuous variable

Description: elec_exp is a continuous variable that refers to total annual household expenditures on electricity and other associated expenditures such as hire of meters, reading of meters and standing charges.

Variable: ngas_exp

Label: Total annual consumption of network/natural gas
Type: Numeric continuous variable
Description: ngas_exp is a continuous variable that refers to total annual household expenditure on town gas and natural gas.

Variable: LPG_exp

Label: Total annual consumption of liquefied gas
Type: Numeric continuous variable
Description: LPG_exp is a continuous variable that refers to total annual household expenditure on LPG that includes butane, propane, "bottled gas" etc.

Variable: gas_exp

Label: Total annual consumption of network/natural and liquefied gas
Type: Numeric continuous variable
Description: gas_exp is a continuous aggregate variable comprised of total annual household expenditures on network/natural gas and liquefied gas (LPG). Due to differences in characteristics and price patterns, two types of gas are recorded as separate variables under gas: 1) Town gas and natural gas (ngas_exp); and 2) LPG (liquefied petroleum gas (LPG_exp): includes butane, propane, "bottled gas", etc.). Associated expenditure such as hire of meters, reading of meters, storage containers, standing charges, etc. are included in the construction of the variable.

Variable: gasoline_exp

Label: Total annual consumption of gasoline
Type: Numeric continuous variable
Description: gasoline_exp is a continuous variable that refers to total annual household expenditure on gasolines. Use mostly in sedan cars and motorcycles.

Variable: diesel_exp

Label: Total annual consumption of diesel
Type: Numeric continuous variable
Description: diesel_exp is a continuous variable that refers to total household expenditure on diesel or gasoil. Mostly use on electricity generators, SUV, Trucks, buses, very few sedan cars use this type of fuel.

Variable: kerosene_exp

Label: Total annual consumption of kerosene
Type: Numeric continuous variable
Description: kerosene_exp is a continuous variable that refers to total annual household expenditure on kerosene.

Variable: othliq_exp

Label: Total annual consumption of other liquid fuels
Type: Numeric continuous variable

Description: othliq_exp is a continuous variable that refers to total annual household expenditure on other liquid fuels such as heating oil, black oil and lighting oil.

Variable: liquid_exp

Label: Total annual consumption of all liquid fuels
Type: Numeric continuous variable
Description: liquid_exp is a continuous aggregate variable comprised of total annual household expenditures on all liquid fuels. Liquid fuels are subcategorized into: gasoline/petrol (gasoline_exp), diesel (diesel_exp), kerosene (kerosene_exp), gasoline (gasoline_exp), and other liquid fuels (othliq_exp). Other liquid fuels category includes all other liquid fuels other than diesel and kerosene. Examples include "heating oil", "black oil" and "lighting oil".

Variable: wood_exp

Label: Total annual consumption of firewood
Type: Numeric continuous variable
Description: wood_exp is a continuous variable that refers to total annual household expenditure on firewood.

Variable: coal_exp

Label: Total annual consumption of coal
Type: Numeric continuous variable
Description: coal_exp is a continuous variable that refers to total annual household expenditure on coal.

Variable: peat_exp

Label: Total annual consumption of peat
Type: Numeric continuous variable
Description: peat_exp is a continuous variable that refers to total annual household expenditure on peat.

Variable: othsol_exp

Label: Total annual consumption of other solid fuels
Type: Numeric continuous variable
Description: othsol_exp is a continuous variable that refers to total annual household expenditure on other solid fuels such as charcoal from wood and agricultural residue.

Variable: solid_exp

Label: Total annual consumption of all solid fuels
Type: Numeric continuous variable
Description: solid_exp is a continuous aggregate variable comprised of total annual household expenditures on all solid fuels. Solid energy is subcategorized into expenditures on coal (coal_exp), firewood (wood_exp) and peat (peat_exp), and other solid fuels (othsol_exp). Other solid fuels category includes all other solid fuels not included in the above three categories. Examples include "pressed dung, corn brans, brushwood", and "other solid".

Variable: othfuel_exp

Label: Total annual consumption of all other fuels
Type: Numeric continuous variable
Description: othfuel_exp is a continuous variable that refers to total annual household expenditure on other fuels that are not captured under othliq_exp and othsol_exp.

Variable: central_exp

Label: Total annual consumption of central heating

Type: Numeric continuous variable

Description: central_exp refers to total annual household expenditure on central heating.

Variable: heating_exp

Label: Total annual consumption of heating

Type: Numeric continuous variable

Description: heating_exp is a continuous aggregate variable comprised of total annual household expenditures on heating. These expenditures can be subcategorized into expenditures on central heating (central_exp) and hot water (hwater_exp). It is worth to note that COICOP narrowly defines heat energy to purchase from district heating plant only, but GMD includes heat energy from building or other sources. Note that expenditure for central heating is frequently combined either with expenditures pm hot water or rent. Hot water is also often combined with cold water. Also note that COICOP categorizes hot water under 4.5.5 Heat energy, while cold water is reflected under 4.4.1 Water supply.

Variable: utl_exp

Label: Total annual consumption of all utilities excluding telecom and other housing

Type: Numeric continuous variable

Description: utl_exp is a continuous aggregate variable comprised of total annual household expenditure on all utilities excluding telecom and other housing expenses. Utilities expenditure in this case is sum of the following variables: electricity (elec_exp), gas (gas_exp), liquid fuels (liquid_exp), solid fuels (solid_exp), central heating (central_exp), water (water_exp), waste (waste_exp) and other fuels (othfuel_exp). Excludes expenditures for other housing (othhousing_exp), fuel for transportation (transfuel_exp), telecommunication services (comm_exp) and tv services (tv_exp).

Variable: dwelmat_exp

Label: Total annual consumption of materials for the maintenance and repair of the dwelling

Type: Numeric continuous variable

Description: dwelmat_exp is a continuous variable that refers to total annual household expenditures on product and materials for maintenance and repair of the dwelling. Products and materials for minor maintenance and repair typically include expenditures on paints and varnishes, renderings, wallpapers, fabric wall coverings, window panes, plaster, cement, putty, wallpaper pastes. Fitted carpets and linoleum (5.1.2); hand tools, door fittings, power sockets, wiring flex and lamp bulbs (5.5.2); brooms, scrubbing brushes, dusting brushes and cleaning products (5.6.1); products, materials and fixtures used for major maintenance and repair (intermediate consumption) or for extension and conversion of the dwelling (capital formation) are excluded.

Variable: dwelsvc_exp

Label: Total annual consumption of services for the maintenance and repair of the dwelling

Type: Numeric continuous variable

Description: dwelsvc_exp is a continuous variable that refers to total annual household expenditures on services for minor maintenance and repair of the dwelling. This variable generally includes expenditures on services of plumbers, electricians, carpenters, glaziers, painters, decorators, floor polishers, etc as well as total value of the service (that is, both the cost of labor and the cost of materials are covered). It excludes separate purchases of materials made by the household with the intention of undertaking the maintenance or repair by themselves (4.3.1); services engaged for major maintenance and repair

(intermediate consumption) or for the extension and conversion of the dwelling (capital formation).

Variable: othhousing_exp

Label: Total annual consumption of dwelling repair/maintenance

Type: Numeric continuous variable

Description: othhousing_exp is a continuous variable that refers to total annual household expenditures on other materials and services for minor maintenance and repair of the dwelling. Use this category for total dwelling repair/maintenance if the survey does not disaggregate expenses into materials and services.

Variable: transfuel_exp

Label: Total annual consumption of fuels for personal transportation

Type: Numeric continuous variable

Description: transfuel_exp is a continuous variable that refers to total annual household expenditures on fuels for personal transportation. According to COICOP, fuels use for transportation purposes are classified under Fuels and lubricants for personal transport equipment (COICOP 7.2.2). COICOP 7.2.2 also includes lubricants, which are excluded from this GMD indicator. If the survey only has variables for gasoline, diesel, or other fuels without explicitly saying that it is for transportation, then we do not include them under transfuel_exp, but under gasoline_exp/diesel_exp/othliq_exp. *Most importantly, these expenditures should NOT be double-counted.*

Variable: landphone_exp

Label: Total annual consumption of landline phone services

Type: Numeric continuous variable

Description: landphone_exp refers to total annual household expenditures on landphone. This includes installation, subscription and service usage fees. Expenditure on equipment are not included.

Variable: cellphone_exp

Label: Total annual consumption expenditures on cellphones

Type: Numeric continuous variable

Description: cellphone_exp is a continuous variable that refers to total annual household expenditures on cellphone. This includes installation, subscription and service usage fees. Expenditure on equipment are not included.

Variable: tel_exp

Label: Total consumption of all telephone services

Type: Numeric continuous variable

Description: tel_exp is a continuous aggregate variable comprised of total annual household expenditures on landline phone (landphone_exp) and cell phone (cellphone_exp) which may include (i) Installation and subscription costs of personal telephone equipment, (ii) telephone calls from a private line or from a public line (public telephone box, post office cabin, etc.); telephone calls from hotels, cafés, restaurants and the like, (iii) hire of telephones, telefax machines, telephone-answering machines and telephone loudspeakers. Expenditures on relevant equipment are not included. Telephone and telefax services (COICOP 8.3.0) are subcategorized into 4 categories: landline phone, cell phone, internet and telefax services.

Variable: internet_exp

Label: Total consumption of internet services
Type: Numeric continuous variable
Description: internet_exp is a continuous variable that refers to total annual household expenditures on information transmission and Internet connection services. This variable also includes installation, subscription, and service usage fees and costs, but excludes consumption for equipment. Telefax services (telefax_exp) includes telegraphy, telex and telefax services, as well as radio-telephony, radio-telegraphy and radiotelex services. Expenditures on relevant equipment are not included.

Variable: telefax_exp
Label: Total consumption of telefax services
Type: Numeric continuous variable
Description: telefax_exp is a continuous variable that refers to total annual household expenditures on telegraphy, telex and telefax services. This includes: radio-telephony, radio-telegraphy and radiotelex services.

Variable: comm_exp
Label: Total consumption of all telecommunication services
Type: Numeric continuous variable
Description: comm_exp is a continuous variable comprised of total annual household expenditures on all telephone and telefax services, including expenditures on landline phone (landphone_exp), cell phone (cellphone_exp), internet (internet_exp) and telefax services (telefax_exp).

Variable: tv_exp
Label: Total consumption of TV broadcasting services
Type: Numeric continuous variable
Description: tv_exp is a continuous variable that refers to total annual household expenditures on television broadcasting services, license fees for television equipment and subscriptions to television networks. This variable is compatible with COICOP 9.4.2 Cultural services but does not include spending on such services as theatres, museums and historic monuments.

Variable: tvintph_exp
Label: Total consumption of tv, internet and telephone
Type: Numeric continuous variable
Description: tvintph_exp is a continuous aggregate variable comprised of total annual household expenditures on internet (internet_exp), telephone (tel_exp) and television broadcasting services (tv_exp).

3.4 Access to Social Amenities

In some surveys this may not be available for each household but will be present in the community survey. The distances and time are to the nearest services from the household irrespective of whether the household uses these services.

All distances and times refer to two-way journeys. Please note that all data for distances and time that are not categorized (continuous) are to the nearest 2 decimal places.

Variable: dispsch

Label: Distance to nearest elementary/primary school (kms)
Type: Numeric continuous variable
Description: One way.
This refers to distance to nearest primary school in kms.
By convention 1 km = 1000 meters; 1 km = 5/8 mile
If roundtrip provided, divide by 2.
If survey question is pre-coded, do not guestimate this into a continuous variable. Leave as missing.

Variable: timpsch
Label: Time taken to nearest elementary/primary school (minutes)
Type: Numeric continuous variable
Description: One way.
This refers to time taken to reach nearest primary school in mins.
By convention 1 hr = 60 min.
If roundtrip provided, divide by 2.
If survey question is pre-coded, do not guestimate this into a continuous variable. Leave as missing.

Variable: disheal
Label: Distance to nearest health facility (kms)
Type: Numeric continuous variable
Description: One way.
This refers to distance to nearest health facility in kms.
By convention 1km = 1000 meters; 1 km = 5/8 mile
If roundtrip provided, divide by 2.
If survey question is pre-coded, do not guestimate this into a continuous variable. Leave as missing.

Variable: timheal
Label: Time taken to nearest health facility (minutes)
Type: Numeric continuous variable
Description: One way
This refers to time taken to reach nearest primary school in mins.
By convention 1hr = 60 min.
If roundtrip provided, divide by 2.
If survey question is pre-coded, do not guestimate this into a continuous variable. Leave as missing.

3.5 Ownership of Durable Assets

Variable: radio
Label: Ownership of radio
Type: Numeric categorical variable
Description: radio is a dummy variable indicating whether the household owns a radio (i.e. radio, radio cassette, and 3-in-1 radio cassette player (radio). Radio ownership does not depend on who owns the radio within the household, nor on its condition. Two categories after harmonization:
0 = No; 1 = Yes

Variable: television
Label: Ownership of television
Type: Numeric categorical variable

Description: tv is a dummy variable indicating whether the household owns a TV set. This includes both color and black and white TVs. TV set ownership does not depend on who owns the TV set within the household, nor on its condition. Two categories after harmonization:
0 = No; 1 = Yes

Variable: television_cable

Label: Ownership of television cable
Type: Numeric categorical variable
Description: television_cable is a dummy variable indicating whether the household owns a cable or dish antenna services. Only for households that reported having a TV (**tv=1**).
Two categories after harmonization:
0 = No; 1 = Yes

Variable: video

Label: Ownership of video
Type: Numeric categorical variable
Description: video is a dummy variable indicating whether the household owns a videocassette player and/or video cassette recorder. Video cassette player ownership does not depend on who owns the player within the household, nor on its condition. Two categories after harmonization:
0 = No; 1 = Yes

Variable: landphone

Label: Ownership of landline (fixed) phone
Type: Numeric categorical variable
Description: landphone is a dummy variable indicating whether the household owns a landline phone. It is generally defined as landline phone, home telephone, or fixed phone. Landline phone ownership does not depend on who owns the phone within the household, nor on its condition.
Two categories after harmonization:
0 = No; 1 = Yes

Variable: cellphone

Label: Ownership of at least one cellular phone
Type: Numeric categorical variable
Description: cellphone is a dummy variable indicating whether anyone in the household owns a cell phone. Cell phone ownership does not depend on who owns the cellphone is within the household, nor on its condition. Two categories after harmonization: 0 = No; 1 = Yes

Variable: phone

Label: Ownership of at least phone
Type: Numeric categorical variable
Description: phone is a dummy variable indicating whether the household owns either a land phone or a cell phone. It should only be coded in cases where the survey does not distinguish between ownership of landline and cell phones. In other cases, it may be coded as missing. Phone ownership does not depend on who owns the phone within the household, nor on its condition. Two categories after harmonization:
0 = No; 1 = Yes

Variable: fridge

Label: Ownership of refrigerator
Type: Numeric categorical variable

Description: fridge is a dummy variable indicating whether the household owns a refrigerator (i.e. refrigerator or freezer). It does not include cooler, icebox or ice chest. Refrigerator ownership does not depend on who owns the asset within the household, nor on its condition. Two categories after harmonization:
0 = No; 1 = Yes

Variable: sewmach

Label: Ownership of sewing machine
Type: Numeric categorical variable
Description: sewmach is a dummy variable indicating whether the household owns a sewing machine. Sewing machine ownership does not depend on who owns the sewing machine within the household, nor on its condition. Two categories after harmonization:
0 = No; 1 = Yes

Variable: washmach

Label: Ownership of washing machine
Type: Numeric categorical variable
Description: washmach is a dummy variable indicating whether the household owns a machine for washing clothes and household linen; but does not include non-electric washing machine. Washing machine ownership does not depend on who owns the asset within the household, nor on its condition. Two categories after harmonization:
0 = No; 1 = Yes

Variable: fan

Label: Ownership of fan
Type: Numeric categorical variable
Description: fan is a dummy variable indicating whether the household owns a fan operated by electricity. Fan ownership does not depend on who owns the asset within the household, nor on its condition. Two categories after harmonization: 0 = No; 1 = Yes

Variable: airconditioner

Label: Ownership of air conditioner
Type: Numeric categorical variable
Description: airconditioner is a dummy variable indicating whether the household owns a central or wall air conditioner. Air conditioner ownership does not depend on who owns the asset within the household, nor on its condition. Two categories after harmonization: 0 = No; 1 = Yes

Variable: computer

Label: Ownership of computer
Type: Numeric categorical variable
Description: computer is a dummy variable indicating whether the household owns a computer, including desktop and laptop computer. Computer ownership does not depend on who owns the computer within the household, nor on its condition. Two categories after harmonization:
0 = No; 1 = Yes

Variable: etablet

Label: Ownership of an electronic tablet
Type: Numeric categorical variable

Description: etablet is a dummy variable indicating the ownership of an electronic tablet. Two categories after harmonization:
0 = No; 1 = Yes

Variable: stove

Label: Ownership of stove
Type: Numeric categorical variable
Description: stove is a dummy variable indicating whether the household owns a stove. Stove generally refers to a portable or fixed apparatus that burns fuel or uses electricity to provide heat for cooking or heating purposes and includes a cooker (stove). Stove ownership does not depend on who owns the asset within the household, nor on its condition. Two categories after harmonization:
0 = No; 1 = Yes

Variable: oxcart

Label: Ownership of animal cart
Type: Numeric categorical variable
Description: oxcart is a dummy variable indicating whether the household owns an animal cart, which is used as a means of transport or a farm tool. Animal cart ownership does not depend on who owns the asset within the household, nor on its condition. Two categories after harmonization:
0 = No; 1 = Yes

Variable: bcycle

Label: Ownership of bicycle
Type: Numeric categorical variable
Description: This dummy variable indicates whether the household owns a bicycle. Note that motored bicycles are classified as motorcycle regardless of motor type. Bicycle ownership does not depend on who owns the asset within the household, nor on its condition. Two categories after harmonization:
0 = No; 1 = Yes

Variable: boat

Label: Ownership of boat
Type: Numeric categorical variable
Description: boat is a dummy variable indicating whether the household owns a boat. Boat ownership does not depend on who owns the asset within the household, nor on its condition. Two categories after harmonization: 0 = No; 1 = Yes

Variable: canoe

Label: Ownership of canoe
Type: Numeric categorical variable
Description: canoe is a dummy variable indicating the ownership of a canoe. Two categories after harmonization: 0 = No; 1 = Yes

Variable: mcycle

Label: Ownership of motorcycle
Type: Numeric categorical variable
Description: mcycle is a dummy variable indicating whether the household owns a motorcycle. Motorcycle refers to an automotive vehicle with two in-line wheels, including motorbike or moped.

Motorcycle ownership does not depend on who owns the asset within the household, nor on its condition. Two categories after harmonization: 0 = No; 1 = Yes

Variable: car

Label: Ownership of private car
Type: Numeric categorical variable
Description: car is a dummy variable indicating whether the household owns a car or truck for household use, excluding commercial vehicle. Car ownership does not depend on who owns the asset within the household, nor on its condition. Two categories after harmonization:
0 = No; 1 = Yes

Variable: Internet

Label: Access to internet inside the house
Type: Numeric categorical variable
Description: internet is a categorical variable indicating whether anyone in the household can use a device that is connected to the internet within the home or have access to internet outside the house. Connection to the Internet can be both wired and wireless and does not depend on who manages it within the household. Four categories after harmonization:
1 = Subscribed in the house
2 = Accessible outside the house (includes internet cafes and smartphones with internet access)
3 = Either (Use this category when the questionnaire does not specify whether the access is in the house or outside the house)
4 = No internet

Variable: ricecook

Label: Ownership of a rice cooker
Type: Numeric categorical variable
Description: ricecook is a dummy variable indicating whether the household owns a rice cooker. Rice cooker ownership does not depend on who owns the asset within the household, nor on its condition. Two categories after harmonization: 0 = No; 1 = Yes

Variable: ewpump

Label: Ownership of an electric water pump
Type: Numeric categorical variable
Description: ewpump is a dummy variable indicating the ownership of an electric water pump. Two categories after harmonization: 0 = No; 1 = Yes

3.6 Household Remittances

Variable: hh_remit

Label: Did household receive any remittances?
Type: Numeric categorical variable
Description: Source of remittances not important here. If HH_REMIT=0 then subsequent questions are null and void. Two categories after harmonization: 0 = No; 1 = Yes

Variable: sex_rmt_1

Label: Sex of the 1st remittance sender
Type: Numeric categorical variable

Description: The order of the sending members is in decreasing order of amount of remittance (remittance includes cash, gifts and food). In some countries, the remittances are by number of transactions, enter each transaction as a unique identifier. This is because one cannot tell if this is the same sender or not. This applies to all questions in this section. Categories : 1 = Male; 0 = Female

Variable: sex_rmt_2

Label: Sex of the 2nd remittance sender
Type: Numeric categorical variable
Description: 1 = Male; 0 = Female

Variable: sex_rmt_3

Label: Sex of the 3rd remittance sender
Type: Numeric categorical variable
Description: 1 = Male; 0 = Female

Variable: relat_rmt_1

Label: Relationship to the household head of the 1st remittance sender
Type: Numeric categorical variable
Description: The order of the sending members is in decreasing order of amount of remittance (remittance includes cash, gifts and food).

2 = Spouse; 3 = Son/daughter
4 = Parents/parents-in-law; 5 = Grandchild
6 = Son-in-law/daughter-in-law; 7 = Other relative
9 = Non-relative

Variable: relat_rmt_2

Label: Relationship to the household head of the 2nd remittance sender
Type: Numeric categorical variable
Description: The order of the sending members is in decreasing order of amount of remittance (remittance includes cash, gifts and food).

2 = Spouse; 3 = Son/daughter
4 = Parents/parents-in-law; 5 = Grandchild
6 = Son-in-law/daughter-in-law; 7 = Other relative
9 = Non-relative

Variable: relat_rmt_3

Label: Relationship to the household head of the 3rd remittance sender
Type: Numeric categorical variable
Description: The order of the sending members is in decreasing order of amount of remittance (remittance includes cash, gifts and food).

2 = Spouse; 3 = Son/daughter
4 = Parents/parents-in-law; 5 = Grandchild
6 = Son-in-law/daughter-in-law; 7 = Other relative
9 = Non-relative

Variable: des_mig_1

Label: Destination of migration of the 1st remittance sending member
Type: Numeric categorical variable

Description: The order of the sending members is in decreasing order of amount of remittance (remittance includes cash, gifts and food).
1 = Capital
2 = Within the country (but not capital)
3 = Abroad

Variable: des_mig_2

Label: Destination of migration of the 2nd remittance sending member
Type: Numeric categorical variable
Description: The order of the sending members is in decreasing order of amount of remittance (remittance includes cash, gifts and food).
1 = Capital
2 = Within the country (but not capital)
3 = Abroad

Variable: des_mig_3

Label: Destination of migration of the 3rd remittance sending member
Type: Numeric categorical variable
Description: The order of the sending members is in decreasing order of amount of remittance (remittance includes cash, gifts and food).
1 = Capital
2 = Within the country (but not capital)
3 = Abroad

Variable: origin_rmt

Label: Origin of the remittance senders
Type: Numeric categorical variable
Description:
1 = Domestic; 2 = Abroad; 3 = Both
Use the following code (if any variable in des_mig_1, des_mig_2 and des_mig_3 is all missing, do not use the following code, edit the code accordingly):

```
gen origin_rmt=1 if inlsit(des_mig_1,1,2)&inlist(des_mig_2,1,2)&inlist(des_mig_3,1,2)
replace origin_rmt=2 if des_mig_1==3&des_mig_2==3&des_mig_3==3
replace origin_rmt=3 if origin_rmt==.
Replace origin_rmt=. If des_mig_1==.&des_mig_2==.&des_mig_3==.
```

Variable: amt_rmt_1

Label: Amount of annual remittance by the 1st remittance sender
Type: Numeric continuous variable
Description: The order of the sending members is in decreasing order of amount of remittance (remittance includes cash, gifts and food).

Variable: amt_rmt_2

Label: Amount of annual remittance by the 2nd remittance sender
Type: Numeric continuous variable
Description: The order of the sending members is in decreasing order of amount of remittance (remittance includes cash, gifts and food).

Variable: amt_rmt_3

Label: Amount of annual remittance by the 3rd remittance sender
Type: Numeric continuous variable
Description: The order of the sending members is in decreasing order of amount of remittance (remittance includes cash, gifts and food).

Variable: amt_rmt_fd
Label: Total amount of annual remittances received in food (annual)
Type: Numeric continuous variable
Description: The total includes the remittances received in the form of food from all remittance senders.

Variable: amt_rmt_oth
Label: Total amount of annual remittances received in other forms (annual)
Type: Numeric continuous variable
Description: The total includes the remittances received in other forms (cash, etc.) from all remittance senders.

4 I Module – Individual-level Variables

This module extracts variables of individuals in the household and contains variables on basic household identification, demographic characteristics, education, migration, and disability.

4.1 Sample and Basic Household Identifier

Variable: country
Label: Country code
Type: string variable
Description: This variable should be created independently from but consistent with other modules.

Variable: year_IHSN
Label: 4-digit year of survey based on IHSN standards
Type: Numeric discrete variable
Description: This variable should be created independently from but consistent with other modules.

Variable: hhno
Label: Household number
Type: Numeric discrete variable
Description: This variable should be created independently from but consistent with other modules.

Variable: hid
Label: Household unique identification
Type: String or numeric variable
Description: This variable should be created independently from but consistent with other modules.

Variable: wta_hh
Label: Household weights
Type: Numeric continuous variable
Description: To obtain household estimates, this is the weight to be used in all computations referring to household-level estimates. This variable cannot be used for poverty estimation. The interpretation is the proportion of households with a certain characteristic is XX%.

4.2 Basic Demographic Characteristics

The file may have different household size when compared to the poverty-level file. Make sure that the regular household members are selected in the same criterion as the Poverty-level file. Secondly, households that do not match the Poverty-level file must be dropped as they do not have the consumption component. All variables are numeric unless specified.

Variable: pid
Label: Individual identifier
Type: string or numeric variable
Description: Uniquely identifies the regular household members in each household. Sequentially numbered from 1 to N (household size). If the PID is a concatenation of HID and person ID, concatenate HID and leave PID only. Check that each household member ID is unique.

```
duplicates tag (hid pid),gen(dup). tab dup
```

Variable: pid_orig

Label: Individual identifier in the raw data

Type: string or numeric, of original data should be kept

Description: This variable is missing if the raw data does not have pid and should be created using other variables (such as region, sector, etc.) . This is the individual ID that was included in the raw data.

Variable: language

Label: language of respondent

Type: String variable

Description: language is a string variable that refers either to the one the respondent normally speaks in his or her present home (usual language) or the language usually spoken in the individual's home in his or her early childhood (mother tongue), or the language that the person commands best (main language). Its classification is country specific. Information on language (including any sign language) should be harmonized for all persons. In the tabulated results, the criterion for determining the language for children not yet able to speak should be clearly indicated. Numeric entries are coded in string format using the following naming convention: "2 – language".

Variable: ageyrs

Label: Age in completed years

Type: Numeric continuous variable

Description: age refers to the interval of time between the date of birth and the data of the survey, expressed in completed solar years. Every effort should be made to determine the precise and accurate age of each person, particularly of children and older persons. Information on age may be secured either by obtaining the date (year, month, and day) of birth or by asking directly for age at the person's last birthday. In addition, in the case of children aged less than or equal to 60 months, variable age should be expressed in decimals***. For example, the age of a respondent who is 6 months old should be recorded as 0.5***. Lastly, if the information on age is not available, it should be coded as missing rather than some other value such as "99" or "999".

If date of birth is provided, derive age and compare with the given recorded age. If age of Household head is missing, use the var=hhagey in the poverty file to replace the missing age of household head only.

For children aged less than 5 years, this is used to interpret child malnutrition and survival data. Check consistency with age in months (AGEM) to get correct age in completed years.

For older surveys, check consistency and maintain AGEYRS.

This can only be done if date of birth and date of interview are provided.

```
gen bday=mdy(month,day,year)
gen iday=mdy(imonth,iday,iyear)
format bday iday %d
gen age = (iday - bday)/365.25
gen ages=trunc(Age)
gen diff=ages-recorded_age
tab diff
```

Variable: agecat

Label: Age intervals (string)

Type: string variable

Description: Country specific categorical variable. It will only be created only when the country does not report the age of the interviewed people but intervals years of their age. Otherwise leave as missing.

```
gen outputvar=""
```

```
qui levelsof inputvar, local(lev)
foreach cc of local lev {
cap loc la_`cc': label(inputvar) `cc'
if !_rc    {
replace outputvar ="`cc'-`la_`cc''" if inputvar ==`cc'
}
}
```

Variable: sex

Label: Sex

Type: Numeric categorical variable

Description: Sex is a dummy variable that specifies the sex – male or female – of an individual within a household. While constructing this variable, it is important to make sure that all relevant values are included. Variable values coded as '98' or other numeric characters should be excluded from the values of the `male' variable. Sex of household member, two categories after harmonization:

1 = Male; 0 = Female

Variable: relathhcs

Label: Relationship to household head (country-specific)

Type: String variable

Description: Country-specific.

For each value label, there should be a space between the hyphen (before and after). Please translate categories to English if necessary.

Code and Variable: Example: "1 - Head"; "2 - Spouse"; "3 – Child"; etc.

```
gen relathhcs=""
qui levelsof inputvar, local(lev)
foreach cc of local lev {
cap loc la_`cc': label(inputvar) `cc'
if !_rc    {
qui replace relathhcs ="`cc' - `la_`cc''" if inputvar ==`cc'
}
}
```

Variable: relathh9

Label: Relationship to household head (9 categories)

Type: Numeric categorical variable

Description: This refers to the relationship of each household member to the household HEAD.

This variable must have one and only one head in each household. Child refers to biological child or adoptive children by either marriage or other reason. Domestic help (servant, guard, cook, baby-sitter among others) refers to a person who is paid for services rendered (cash or in-kind e.g. training skills, board and lodging) even if they are related to the head of household. Paying boarder is someone who pays the household for room and/or board. None relative include friends living in household regularly. In cases where head is missing or a migrant, we assign spouse as the head of the household. If spouse is also not available, then we will use oldest member of the household as the head and recode all the relations to head accordingly. Use relathhcs to derive this variable after the edits. If all categories are not present in the questionnaire, leave this variable as missing

1 = Head; 2 = Spouse

3 = Child 4 = Parents/parents-in-law

5 = Grandchild 6 = Son-in-law/daughter-in-law

7 = Other relative 8 = Domestic help/paying boarder

9 = None relative

Variable: relathh6

Label: Relationship to household head (6 categories)

Type: Numeric categorical variable

Description: This refers to the relationship of each household member to the household HEAD. Must have one and only one head in each household. Other includes grandchild, in-laws, etc. Non-relative includes domestic help, paying boarder, etc.

1 = Head
2 = Spouse
3 = Child
4 = Parents
5 = Other relative
6 = Non-relative

```
recode relathh9 (1=1) (2=2) (3=3) (4=4) (5/7=5) (8/9=6), gen(relathh6)
```

Variable: marital6

Label: Marital status (6 categories)

Type: Numeric categorical variable

Description: Polygamous unions exclude relationships that are not officially recognized such as mistresses, concubines. Check for consistency in married unions. Marital status for couples must be identical. Do not derive polygamous unions if survey does not ask. Leave variable as missing.
If marital asked for persons only above 12 years, one can confidently guestimate that the children are "Never married". If all categories are not present in the questionnaire, leave this variable as missing.

1 = Married monogamous
2 = Married polygamous
3 = Never married
4 = Living together
5 = Divorced/separated
6 = Widowed

Variable: marital5

Label: Marital status (5 categories)

Type: Numeric categorical variable

Description: marital5 is a categorical variable that refers to the personal status of each individual in relation to the marriage laws or customs of the country. This variable should include at least the following: (a) married; (b) never married; (c) living together; (d) divorced/separated; (e) widowed. In some countries, category (a) may require a subcategory of persons who are contractually married but not yet living as man and wife. In all countries, category (d) should comprise both the legally and the de facto separated, who may be shown as separate subcategories if desired. The marital variable should not be imputed but rather calculated only for those to whom the question was asked (in other words, the youngest age at which information is collected may differ depending on the survey).

The consistency between age and marital5 needs to be cross-checked. In most countries, there are also likely to be persons who were permitted to marry below the legal minimum age because of special

circumstances. To permit international comparisons of data on marital status, however, any tabulations of marital status not cross-classified by exact age should at least distinguish between persons under 15 years of age and over. If it is not possible to distinguish between married and living together, then it should be assumed that the individual is married. Variable values coded as '98' or other numeric characters should be excluded from the values of the 'marital5' variable.

1 = Married
2 = Never Married
3 = Living together
4 = Divorced/Separated
5 = Widowed

```
recode marital6 (1 2=1) (3=2) (4=3) (5=4) (6=5), gen(marital5)
tab marital6 marital5
```

Variable: sp_pres

Label: Spouse of household head living in household

Type: Numeric categorical variable

Description: Code based on a question that asks whether the household head spouse lives in the household. Otherwise leave as missing. Only for marital5 = 1 or 3. DO NOT TRY TO DEDUCE FROM HOUSEHOLD MEMBERSHIP. However, under some special circumstances, a couple may be divorced/separated but living in the same household (dwelling unit) but in separate rooms. In this instance, sp_pres=1. Categories after harmonization: 0 = No; 1 = Yes

4.3 Literacy and Education

Variable: literacy

Label: Literacy status

Type: Numeric categorical variable

Description: For individuals aged 5 and above only. Value must be missing for all others.

Literacy: Is the ability to both read and write with understanding, a short simple statement on his/her everyday life in any language. It will be useful to align measurements of literacy with this given standard international definition.

Be careful while coding 1; one must be able to both read and write. If a person can either read or write, he/she will be considered illiterate (LITERACY=0). It can be assumed with some degree of accuracy that if respondent has secondary level and above of education, then must be literate.

Also, persons with over 5 years of primary can be assumed literate. Can be programmed with EDUCYRS if literacy is missing for some members.

1 = Yes, can read and write
0 = No, cannot read or write

Variable: ed_mod_age

Label: Education module application age (country-specific)

Type: Numeric categorical variable

Description: Minimum age for which education section is applied in country. The questionnaire and/or manual specifies this. For this reason, the lower age cutoff at which information is collected will vary from country to country.

Variable: everattd

Label: Ever attended school
Type: Numeric categorical variable
Description: Value must be missing for individuals less than the required age (ed_mod_age).
Depends on how school attendance is defined in a country. Example, in some countries, a criterion is placed to decide if ever attended school is valid or not and is determined by number of weeks or months or school term in attendance. Does not require to have completed any level of education.
Indirect derivation if not collected by survey would be to program EDUCAT10 and ATSCHOOL. If ATSCHOOL=1 then ever attended=1. If EDUCAT10>=3 and EDUCAT10<=9, ever attended = 1.
Two categories after harmonization: 0 = No; 1 = Yes

Variable: educat10

Label: Highest level of education completed (10 categories)
Type: Numeric categorical variable
Description: Value must be missing for individuals less than the required age (ed_mod_age).
If a person is currently enrolled in the highest year of education, then his/her level of education completed should be determined by minus one year. For example, if a person is currently enrolled in P6, then his/her highest level completed should be coded as 1 (Pre-school/ Primary, not completed).
Individuals enrolled in University level are coded as 8 (University and higher) regardless of whether completed or not. Other refers to level of education not defined by the above codes. This may refer to level of education not explicitly defined e.g. person attending a village polytechnic, yet level reached not stated. This classification should be documented whenever possible.

If Koranic school teaches formal curricula then it will be classified under formal education, then code appropriately.

Koranic schools that teach Islamic knowledge with only (a) basic recitation or (b) recitation and Arabic writing or hafeez (memorization and Arabic fluency) are not mainstream formal schools. Code as "Other"
If education level is missing for any member, do not try to impute but leave it as MISSING.
If all categories are not present in the questionnaire, leave this variable as missing.

10 Categories after harmonization:
1 = No education
2 = Preschool
3 = Primary incomplete
4 = Primary complete but less than completed lower secondary
5 = Completed lower secondary (or post-primary vocational education) but less than completed upper secondary
6 = Completed upper secondary (or extended vocational/technical education)
7 = Post-secondary but not university
8 = University and higher
9 = Formal adult education or literacy program
10 = Other

Variable: educat7

Label: Highest level of education completed (7 categories)
Type: Numeric categorical variable
Description: Value must be missing for individuals less than the required age (ed_mod_age).
Primary complete implies that one completed the stipulated primary education by undertaking an exam or test. Secondary complete implies that one completed the stipulated secondary education by undertaking an exam or test.

Post-secondary technical education level refers to any higher education after successfully completing secondary level of education such as higher professional schooling, college, etc.
University and higher education level refer to undergraduate and higher.
If education level is missing, do not try to impute but leave it as MISSING.
If all categories are not present in the questionnaire, leave this variable as missing.

1 = No education
2 = Primary incomplete
3 = Primary complete
4 = Secondary incomplete
5 = Secondary complete
6 = Post-secondary but not university
7 = University (complete or incomplete)

Variable: educat5

Label: Highest level of education completed (5 categories)
Type: Numeric categorical variable
Description: Value must be missing for individuals less than the required age (ed_mod_age).
If education level is missing, do not try to impute but leave it as MISSING.
If all categories are not present in the questionnaire, leave this variable as missing.

1 = No education
2 = Primary incomplete
3 = Primary complete but Secondary incomplete
4 = Secondary complete
5 = Tertiary/post-secondary (complete or incomplete)

Can be programmed from educat7.

```
recode educat7 (3 4=3) (5=4) (6 7=5), gen(educat5)
tab ageyrs educat5
```

Variable: educat4

Label: Highest level of education completed (4 categories)
Type: Numeric categorical variable
Description: Value must be missing for individuals less than the required age (ed_mod_age).
No education includes people in pre-school and never attended. Pre-school definition is country-specific. This may include baby class, kindergarten and nursery school among others. This is the level before joining the regular stipulated primary level education cycle. At the minimum, educat4 must be available for all countries. If education level is missing, do not try to impute but leave it as MISSING.
4 categories after harmonization:
1 = No education
2 = Primary (complete or incomplete)
3 = Secondary (complete or incomplete)
4 = Tertiary (complete or incomplete)
Can be programmed from educat7.

```
recode educat7 (2 3=2) (4 5=3) (6 7=4),gen(educat4)
tab ageyrs educat4
```

Variable: educat_ISCED

Label: ISCED education categories (highest level enrolled in or completed)

Type: Numeric categorical variable

Description: These are the UNESCO ISCED 2011 education categories. Please note that we use the highest level enrolled in or completed. For example, if you are enrolled in primary education, you should get category 2 even if you have not completed primary yet or never will.

Check this link for country ISCED Mappings 9

Post-secondary non-tertiary education may be referred in many ways depending on country. However, these are typically vocational programmes that prepare one for the labor market such as technician diploma, electrician diploma.

1 = Early childhood education
2 = Primary education
3 = Lower secondary education
4 = Upper secondary education
5 = Post-secondary non-tertiary education
6 = Short-cycle tertiary education
7 = Bachelor's or equivalent level
8 = Master's or equivalent level
9 = Doctoral or equivalent level

Variable: primarycomp

Label: Primary school completion

Type: Numeric categorical variable

Description: Value must be missing for other individual less than the required age (ed_mod_age).

One can assume with a degree of certainty these conditions qualify primary-school completion:

EDUCAT10>=4 & EDUCAT10<=8; EDUCAT7>=3 & EDUCAT&<=7; EDUCAT5>=3 & EDUCAT5<=5

0 = No; 1 = Yes

Variable: educyrs

Label: Years of completed education

0 = Pre-school; 1 = Grade 1; 2 = Grade 2 …

Type: Numeric categorical variable

Description: It is constructed only if the survey asked for the number of years of education or highest grade level completed; otherwise, the values are constructed as missing.

Value must be missing for other individual less than the required age (ed_mod_age). If grade level not listed, leave EDUCYRS=. For individuals who are currently enrolled in school, their years of education completed correspond to the class currently attending minus one. For individuals who are not currently enrolled in school, the years of completed education corresponds to the highest level of education completed.

The years of education that each grade corresponds to, varies by country, for example - some countries may have 5 or 6 years of primary school, 3 years of lower-secondary school, while other countries may have 4 years of primary school and 4 years of lower-secondary school. Refer to the UNESCO ISCED mappings.

For higher education, the grades/years may not have been asked explicitly. In such cases, the variable should be constructed based on the following assumptions: -

- If the individual has completed the tertiary education specified, add to years of completed education - 4 years for BA/BSc, 6 years for MA/MSc, and 8 Years for PhD after the completion of secondary education.
- If the individual has not completed tertiary education or completion cannot be ascertained, add to years of completed education – 2 years for BA/BSc, 5 years for MA/MSc, and 7 years for PhD.

The variable does not take into account the actual number of years required to reach this grade level. In other words, first grade repeated three times only counts as 1 year of completed education.

Variable: atschool

Label: Currently enrolled in or attending school
Type: Numeric categorical variable
Description: Value must be missing for individuals less than the required age (ed_mod_age).
Use the question that asks for current attendance.
If such a question is missing, use the question that explicitly asks for enrollment over the past 12 months. In such surveys, record this in the comments.
Code as 0 if EVERATTD=0.
Two categories after harmonization: 0 = No; 1 = Yes

Variable: atschltyp

Label: Type of school currently enrolled/attending
Type: Numeric categorical variable
Description: Value must be missing for individuals less than the required age (ed_mod_age).
Code only for individuals currently attending school (ATSCHOOL=1).
Public includes fully government owned as well as semi-public owned.
Private are facilities run by non-governmental organizations (e.g. NGOs, religious institutions) or by private entities.
Other refers to schools that cannot be categorized by the above such as community schools which cannot be easily classified if run by either government or private.
Three categories after harmonization: 1 = Public; 2 = Private; 9 = Other

Variable: atslevattd

Label: Level of schooling currently enrolled/attending
Type: Numeric categorical variable
Description: Value must be missing for individuals less than the required age (ed_mod_age).
See EDUCAT10 for definition.
Check for consistency between EDUCAT10. That is EDUCAT10 cannot be university yet current level primary.

1 = Preschool
2 = Primary
3 = Secondary
4 = Post-secondary but not university
5 = University and higher
6 = Formal adult education or literacy program
9 = Other

4.4 Migration

Variable: rb_mod_age

Label: Migration module application age (country-specific)
Type: Numeric discrete variable
Description: Minimum age for which migration is applied.
For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country.

Variable: rbirth

Label: Was member born in this country?
Type: Numeric categorical variable
Description: Value must be missing for individuals less than the required age (rb_mod_age).
0 = No; 1 = Yes

Variable: rbirth_ctry

Label: In what country was member born?
Type: String variable
Description: Value must be missing for individuals less than the required age (rb_mod_age).
Only if RBIRTH=0.
If born outside country, enter 3-digit ISO country code (see Annex X).
Several codes added for use if country no specified.
“Other Africa”
“Other Europe”
“Other America”
“Other (unspecified)”

Variable: rbirthreg

Label: Was person born in this region?
Type: Numeric categorical variable
Description: Value must be missing for individuals less than the required age (rb_mod_age).
0 = No; 1 = Yes

Variable: rbirth_reg

Label Region of birth
Type: String variable
Description: Value must be missing for individuals less than the required age (rb_mod_age).
Only if RBIRTH_REG==0
Use survey region codes. Must entered as “1 – region 1 name”, “2 – region 2 name”, etc.

Variable: rbirth_prevref

Label: Reference time for previous residence
Type: String variable
Description: Indicates the time reference of the question about migration (or place of residence).
For example, RBIRTH_PREV_REF=5, means that the question asks about place of residence 5 years ago.

Variable: rbirthprev

Label: Ever lived in a previous residence than the current one?
Type: Numeric categorical variable
Description: Value must be missing for individuals less than the required age (rb_mod_age).
If person lived in several places, only the most recent should be recorded here.
1 = Yes, within county
2 = Yes, outside country
3 = No

Variable: rbirth_prev

Label: Region of previous residence
Type: String variable
Description: Value must be missing for individuals less than the required age (rb_mod_age).
Only if RBIRTHPREV==1. If survey asks by area of residence, leave this variable as missing.
Code using region codes of survey, must entered as "1 - region name", etc.
Code using region codes of survey, must entered as "1 - region name", etc.

Variable: ymove

Label: Year individual moved to current location
Type: Numeric continuous variable
Description: Value must be missing for individuals less than the required age (rb_mod_age).
Indicates year of most recent move to RBIRTH_PREV.

4.5 Disability

Variable: eye_dsablty

Label: Eye Disability
Type: Numeric categorical variable
Description: eye_dsablty is a Numeric variable that indicates whether an individual has any difficulty in seeing, even when wearing glasses. Two categories after harmonization:
1 = No – no difficulty
2 = Yes – some difficulty
3 = Yes – a lot of difficulty
4 = Cannot do at all

Variable: hear_dsablty

Label: Hear Disability
Type: Numeric categorical variable
Description: hear_dsablty is a Numeric variable that indicates whether an individual has any difficulty in hearing even when using a hearing aid.
1 = No – no difficulty
2 = Yes – some difficulty
3 = Yes – a lot of difficulty
4 = Cannot do at all

Variable: walk_dsablty

Label: walk Disability

Type: Numeric categorical variable

Description: walk_dsablty is a Numeric variable that indicates whether an individual has any difficulty in walking or climbing steps.

1 = No – no difficulty
2 = Yes – some difficulty
3 = Yes – a lot of difficulty
4 = Cannot do at all

Variable: conc_dsord

Label: Concentration Disorder

Type: Numeric categorical variable

Description: conc_dsord is a Numeric variable that indicates whether an individual has any difficulty concentrating or remembering

1 = No – no difficulty
2 = Yes – some difficulty
3 = Yes – a lot of difficulty
4 = Cannot do at all

Variable: slfcre_dsablty

Label: Self-care Disability

Type: Numeric categorical variable

Description: slfcre_dsablty is a Numeric variable that indicates whether an individual has any difficulty with self-care such as washing all over or dressing.

1 = No – no difficulty
2 = Yes – some difficulty
3 = Yes – a lot of difficulty
4 = Cannot do at all

Variable: comm_dsablty

Label: Communication Disability

Type: Numeric categorical variable

Description: comm_dsablty is a Numeric variable that indicates whether an individual has any difficulty communicating or understanding usual (customary) language.

1 = No – no difficulty
2 = Yes – some difficulty
3 = Yes – a lot of difficulty
4 = Cannot do at all

5 L Module – Labor Variables

To the extent possible, variables in this module should be generated independently from the I module. If necessary, you can copy code to generate the basic demographic variables. Gross wages should be used when available and net wages only when gross wages are not available. This is done to make it easy to compare wage earnings between formal and informal sectors.

5.1 Sample and Basic Household Identifier

Variable: country
Label: Country code
Type: String variable
Description: This variable should be created independently from but consistent with other modules.

Variable: year_IHSN
Label: 4-digit year of survey based on IHSN standards
Type: Numeric discrete variable
Description: This variable should be created independently from but consistent with other modules.

Variable: hhno
Label: Household number
Type: Numeric discrete variable
Description: This variable should be created independently from but consistent with other modules.

Variable: hid
Label: Household unique identification
Type: String or numeric variable
Description: This variable should be created independently from but consistent with other modules.

Variable: wta_hh
Label: Household weights
Type: Numeric continuous variable
Description: To obtain household estimates, this is the weight to be used in all computations referring to household-level estimates. This variable cannot be used for poverty estimation. The interpretation is the proportion of households with a certain characteristic is XX%.

5.2 Labor status, 7-day reference period

Variable: pid
Label: Individual identification
Type: String or numeric variable
Description: See I module for details on this variable

Variable: ageyrs
Label: Age in completed years
Type: Numeric continuous variable
Description: See I module for details on this variable

Variable: minlaborage

Label: Labor module application age (7-day ref period)

Type: Numeric discrete variable

Description: This is the lowest age for which the labor module is implemented in the survey or the minimum working age in the country. For this reason, the lower age cutoff at which information is collected will vary from country to country.

Variable: lstatus

Label: Labor status (7-day ref period)

Type: Numeric categorical variable

Description: lstatus is an individual's labor status in the last 7 days. The value must be missing for individuals less than the required age (minlaborage).

Three categories are used after harmonization:

1 = Employed; 2 = Unemployed; 3 = Not-in-labor force

All persons are considered active in the labor force if they presently have a job (formal or informal, i.e., employed) or do not have a job but are actively seeking work (i.e., unemployed).

1 = Employed

Employed is defined as anyone who worked during the last 7 days or reference week, regardless of whether the employment was formal or informal, paid or unpaid, for a minimum of 1 hour. Individuals who had a job, but for any reason did not work in the last 7 days are considered employed.

2 = Unemployed

A person is defined as unemployed if he or she is, presently not working but is actively seeking a job. The formal definition of unemployed usually includes being 'able to accept a job.' This last question was asked in a minority of surveys and is, thus, not incorporated in the present definition. A person presently not working but waiting for the start of a new job is considered unemployed.

3 = Not-in-labor force

A person is defined as not-in-labor force if he or she is, presently not working and it is not actively seeking a job during the last 7 days or reference week.

Variable: nlfreason

Label: Reason not in the labor force (7-day ref period)

Type: Numeric categorical variable

Description: nlfreason is the reason an individual was not in the labor force in the last 7 days. This variable is constructed for all those who are not presently employed and are not looking for work (lstatus=3) and missing otherwise.

Five categories after harmonization:

1= Student (a person currently studying.)

2= Housewife (a person who takes care of the house, older people, or children)

3= Retired

4 = Disabled (a person who cannot work due to physical conditions)

5 = Other (a person does not work for any other reason)

Fill this information for all people interviewed in the labor section of the questionnaire regardless of their age.

Variable: unempldur_l

Label: Unemployment duration (months) lower bracket (7-day ref period)

Type: Numeric continuous variable

Description: unempldur_l is a continuous variable specifying the duration of unemployment in months (lower bracket).

The variable is constructed for all unemployed persons (lstatus=2, otherwise missing). If it is specified as continuous in the survey, it records the numbers of months in unemployment. If the variable is categorical it records the lower boundary of the bracket.

Missing values are allowed for everyone who is not unemployed.

Variable: unempldur_u

Label: Unemployment duration (months) lower bracket (7-day ref period)

Type: Numeric continuous variable

Description: unempldur_u is a continuous variable specifying the duration of unemployment in months (upper bracket).

The variable is constructed for all unemployed persons (lstatus=2, otherwise missing). If it is specified as continuous in the survey, it records the numbers of months in unemployment. If the variable is categorical it records the upper boundary of the bracket. If the right bracket is open a missing value should be inputted.

Missing values are allowed for everyone who is not unemployed.

If the duration of unemployment is not reported as a range, but as continuous variables, the unempldur_l and unempldur_u variables will have the same value. If the high range is open-ended the unempldur_u variable will be missing.

5.3 Primary Employment, 7-day reference period

Variable: empstat

Label: Employment status, primary job (7-day ref period)

Type: Numeric categorical variable

Description: empstat is a categorical variable that specifies the main employment status in the last 7 days of any individual with a job (lstatus=1) and is missing otherwise. The variable is constructed for all individuals that respond to this question, even if they are below the working age. For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country.

The definitions are taken from the International Labor Organization's Classification of Status in Employment with some revisions to consider the data available.

Five categories after harmonization:

1 = Paid Employee; 2 = Non-Paid Employee

3 = Employer ; 4 = Self-employed

5 = Other, workers not classifiable by status

1 = Paid Employee

Paid employee includes anyone whose basic remuneration is not directly dependent on the revenue of the unit they work for, typically remunerated by wages and salaries but may be paid for piece work or in-kind. The 'continuous' criteria used in the ILO definition is not used here as data are often absent and due to country specificity.

2 = Non-Paid Employee
Non-paid employee includes contributing family workers who hold a self-employment job in a market-oriented establishment operated by a related person living in the same households who cannot be regarded as a partner because of their degree of commitment to the operation of the establishment, in terms of working time or other factors, is not at a level comparable to that of the head of the establishment. All apprentices should be mapped as 'non-paid employee'
3 = Employer
An employer is a business owner (whether alone or in partnership) with employees. If the only people working in the business are the owner and contributing family workers, the person is not considered an employer (as has no employees) and is, instead classified as self-employed.
4 = Self-employed
Own account or self-employment includes jobs where remuneration is directly dependent from the goods and service produced (where home consumption is considered to be part of the profits) and where one has not engaged any permanent employees to work for them on a continuous basis during the reference period.
Members of producers' cooperatives are workers who hold a self-employment job in a cooperative producing goods and services, in which each member takes part on an equal footing with other members in determining the organization of production, sales and/or other work of the establishment, the investments and the distribution of the proceeds of the establishment amongst the members.
5 = Other, workers not classifiable by status
Other, workers not classifiable by status include those for whom insufficient relevant information is available and/or who cannot be included in any of the above categories.

Variable: ocusec

Label: Sector of activity, primary job (7-day ref period)
Type: Numeric categorical variable
Description: ocusec is a categorical variable that specifies the sector of activity in the last 7 days. It classifies the main job's sector of activity of any individual with a job (lstatus=1) and is missing otherwise. The variable is constructed for all individuals that respond to this question, even if they are below the working age.
Four categories after harmonization:
1 = Public sector, Central Government, Army (including armed forces)
2 = Private, NGO
3 = State-owned
4 = Public or State-owned, but cannot distinguish

1.Public Sector, Central Government, Army (including armed forces)
Public sector is the part of economy run by the government.
2 = Private, NGO
Private sector is that part of the economy which is both run for private profit and is not controlled by the state, it also includes non-governmental organizations
3 = State-owned enterprises
State-owned includes para-state firms and all others in which the government has control (participation over 50%).
4 = Public or State-owned, but cannot distinguish
Select this option is the questionnaire does not ask for State-owned enterprises, and only for Public sector.
Notes: Do not code basis of occupation (ISCO) or industry (ISIC) codes.

Variable: industry_orig

Label: Original industry code, primary job (7-day ref period)

Type: String variable

Description: industry_orig is a string variable that specifies the original industry codes in the last 7 days for the main job provided in the survey (the actual question) and should correspond to whatever is in the original file with no recoding. The variable is constructed for all individuals that respond to this question, even if they are below the working age. It classifies the main job of any individual with a job (lstatus=1) and is missing otherwise

Variable: industrycat10

Label: 1 digit industry classification, primary job (7-day ref period)

Type: Numeric categorical variable

Description: industrycat10 is a categorical variable that specifies the 1-digit industry classification in the last 7 days for the main job of any individual with a job (lstatus=1) and is missing otherwise. The variable is constructed for all individuals that respond to this question, even if they are below the working age. The codes for the main job are given here based on the UN International Standard Industrial Classification. It classifies the main job of any individual with a job (lstatus=1) and is missing otherwise

Ten categories after harmonization:

1 = Agriculture, Hunting, Fishing, etc.
2 = Mining
3 = Manufacturing
4 = Public Utility Services
5 = Construction
6 = Commerce
7 = Transport and Communications
8 = Financial and Business Services
9 = Public Administration
10 = Other Services, Unspecified

Notes:

In the case of different classifications (former Soviet Union republics, for example), recoding has been done to best match the ISIC codes. Category 10 is also assigned for unspecified categories or items.

If all 10 categories cannot be identified in the questionnaire create this variable as missing and proceed to create industrycat4.

Variable: industrycat4

Label: 4-category industry classification, primary job (7-day ref period)

Type: Numeric categorical variable

Description: industrycat4 is a categorical variable that specifies the 1-digit industry classification in the last 7 days for the main job for Broad Economic Activities. This variable is either created directly from the data (if industry classification does not exist for ten categories) or created from industrycat10.

Four categories after harmonization:

1 = Agriculture; 2= Industry; 3 = Services; 4 = Other

Variable: occup_orig

Label: Original occupational classification, primary job (7-day ref period)

Type: String variable

Description: occup_orig is a string variable that specifies the original occupation code in the last 7 days for the main job. This variable corresponds to whatever is in the original file with no recoding.

Variable: occup

Label: 1 digit occupational classification, primary job (7-day ref period)

Type: Numeric categorical variable

Description: occup is a categorical variable that specifies the 1-digit occupational classification for the main job in the last 7 days of any individual with a job (lstatus=1) and is missing otherwise. The variable is constructed for all individuals that respond to this question, even if they are below the working age. For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country. Most surveys collect detailed information and then code it, without keeping the original data, no attempt has been made to correct or check the original coding. The classification is based on the International Standard Classification of Occupations (ISCO). It classifies the main job of any individual with a job (lstatus=1) and is missing otherwise. Eleven categories after harmonization:

1 = Managers
2 = Professionals
3 = Technicians and associate professionals
4 = Clerical support workers
5 = Service and sales workers
6 = Skilled agricultural, forestry and fishery workers
7 = Craft and related trades workers
8 = Plant and machine operators, and assemblers
9 = Elementary occupations
10 = Armed forces occupations
99 = Other/unspecified

Variable: wage_nc

Label: Last wage payment, primary job, excl. bonuses, etc. (7-day ref period)

Type: Numeric continuous variable

Description: wage_nc is a continuous variable that specifies the last wage payment in local currency of any individual (lstatus=1 & empstat=1) in its primary occupation at the reference period reported in the survey and it is missing otherwise. The wage should come from the main job, in other words, the job that the person dedicated most time in the week preceding the survey. This excludes tips, bonuses, other compensation such as dwellings or clothes, and other payments. The variable is constructed for all persons administered this module in each questionnaire. For this reason, the lower age cutoff (and perhaps upper age cutoff) will vary from country to country. Notes:

- For all those with self-employment or owners of own businesses, this should be net revenues (net of all costs EXCEPT for tax payments) or the amount of salary taken from the business. Due to the almost complete lack of information on taxes, the wage from main job is NOT net of taxes.
- By definition, non-paid employees (empstat=2) should have wage=0.
- The reference period of the wage_nc will be recorded in the unitwage variable.

Variable: unitwage

Label: Time unit of last wages payment, primary job (7-day ref period)

Type: Numeric categorical variable

Description: unitwage is a categorical variable that specifies the time reference for the wage_nc variable. It specifies the time unit measurement for the wages of any individual (lstatus=1 & empstat=1) and it is missing otherwise. Acceptable values include:

1 = Daily
2 = Weekly
3 = Every two weeks
4 = Every two months
5 = Monthly
6 = Quarterly
7 = Every six months
8 = Annually

9 = Hourly 10 = Other

Variable: whours

Label: Hours of work in last week, primary job (7-day ref period)

Type: Numeric continuous variable

Description: whours is a continuous variable that specifies the hours of work last week for the main job of any individual with a job (lstatus=1) and is missing otherwise. The main job defined as that occupation that the person dedicated more time to over the past week. The variable is constructed for all persons administered this module in each questionnaire. Notes:

- If the respondent was absent from the job in the week preceding the survey due to holidays, vacation, or sick leave, then record the time worked in the previous 7 days that the person worked.
- Sometimes the questions are phrased as, "on average, how many hours a week do you work?".
- For individuals who only give information on how many hours they work per day and no information on number of days worked a week, multiply the hours by 5 days.
- In the case of a question that has hours worked per month, divide by 4.3 to get weekly hours.

Variable: wmonths

Label: Months worked in the last 12 months, primary job (7-day ref period)

Type: Numeric continuous variable

Description: wmonths is a continuous variable that specifies the number of months worked in the last 12 months for the main job of any individual with a job (lstatus=1) and is missing otherwise. The main job is defined as that occupation that the person dedicated more time to over the past week. The variable is constructed for all persons administered this module in each questionnaire.

Variable: wage_total

Label: Annualized total wage, primary job (7-day ref period)

Type: Numeric continuous variable

Description: wage_total is a continuous variable that specifies the annualized wage payment (regular wage plus bonuses, in-kind, compensation, etc.) for the primary occupation in local currency of any individual (lstatus=1 & empstat=1) and is missing otherwise. The wage should come from the main job, in other words, the job that the person dedicated most time in the week preceding the survey. This wage includes tips, compensations such as bonuses, dwellings or clothes, and other payments. wage_total should be equal to wage_nc in case there are no bonuses, tips etc. offered as part of the job. The variable is constructed for all persons administered this module in each questionnaire.

The annualization of the wage_total should consider the number of months/weeks the persons have been working and receiving this income. You should not assume the person has been working the whole year.

Example: Creation of wage_total when there are no bonuses nor other compensations

```
gen double wage_total=.
replace wage_total=(wage_nc*5*4.3)*wmonths if unitwage==1 //Wage daily
replace wage_total=(wage_nc*4.3)*wmonths if unitwage==2 //Wage weekly
replace wage_total=(wage_nc*2.15)*wmonths if unitwage==3 //Wage every 2 weeks
replace wage_total=(wage_nc)/2*wmonths if unitwage==4 //Wage every 2 months
replace wage_total=( wage_nc)*wmonths if unitwage==5 //Wage monthly
replace wage_total=( wage_nc)/3*wmonths if unitwage==6 //Wage quarterly
replace wage_total=( wage_nc)/6*wmonths if unitwage==7 //Wage every six months
replace wage_total= wage_nc/12*wmonths  if unitwage==8 //Wage annual
```

```
replace wage_total=(wage_nc*whours*4.3)*wmonths if unitwage==9 //Wage hourly
```

Variable: contract

Label: Contract (7-day ref period)
Type: Numeric categorical variable
Description: contract is a dummy variable that classifies the contract status (yes/no) of any individual with a job (lstatus=1) and is missing otherwise. It indicates whether a person has a signed (formal) contract, regardless of duration. The variable is constructed for all persons administered this module in each questionnaire. Two categories after harmonization:
0 = No; 1 = Yes

Variable: healthins

Label: Health insurance (7-day ref period)
Type: Numeric categorical variable
Description: healthins is a dummy variable that classifies the health insurance status (yes/no) of any individual with a job (lstatus=1) and is missing otherwise. Variable is constructed for all persons administered this module in each questionnaire. However, this variable is only constructed if there is an explicit question about health insurance provided by the job. Two categories after harmonization:
0 = No; 1 = Yes

Variable: socialsec

Label: Social security (7-day ref period)
Type: Numeric categorical variable
Description: socialsec is a dummy variable that classifies the social security status (yes/no) of any individual with a job (lstatus=1) and is missing otherwise. Variable is constructed for all persons administered this module in each questionnaire. For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country. However, this variable is only constructed if there is an explicit question about pension plans or social security. Two categories after harmonization: 0 = No; 1 = Yes

Variable: union

Label: Union membership (7-day ref period)
Type: Numeric categorical variable
Description: union is a dummy variable that classifies the union membership status (yes/no) of any individual with a job (lstatus=1) and is missing otherwise. Variable is constructed for all persons administered this module in each questionnaire. For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country. However, this variable is only constructed if there is an explicit question about trade unions. Two categories after harmonization:
0 = No; 1 = Yes

Variable: firmsize_l

Label: Firm size (lower bracket), primary job (7-day ref period)
Type: Numeric continuous variable
Description: firmsize_l specifies the lower bracket of the firm size. The variable is constructed for all persons who are employed in the last 7 days for the main job. If it is continuous, it records the number of

people working for the same employer. If the variable is categorical, it records the lower boundary of the bracket.

Variable: firmsize_u

Label: Firm size (upper bracket), primary job (7-day ref period)
Type: Numeric continuous variable
Description: firmsize_u specifies the upper bracket of the firm size. The variable is constructed for all persons who are employed in the last 7 days for the main job. If it is continuous, it records the number of people working for the same employer. If the variable is categorical, it records the upper boundary of the bracket. If the right bracket is open, this variable should be missing.

5.4 Secondary Employment, 7-day reference period

Variable: empstat_2

Label: Employment status, secondary job (7-day ref period)
Type: Numeric categorical variable
Description: empstat_2 is a categorical variable that specifies employment status of the secondary job with reference period of last 7 days of any individual with a job (lstatus=1) and is missing otherwise. The variable is constructed for all individuals that respond to this question, even if they are below the working age. For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country.
The definitions are taken from the International Labor Organization's Classification of Status in Employment with some revisions to consider the data available.
Five categories after harmonization:
1 = Paid Employee 2 = Non-Paid Employee
3 = Employer 4 = Self-employed
5 = Other, workers not classifiable by status

Variable: ocusec_2

Label: Sector of activity, secondary job (7-day ref period)
Type: Numeric categorical variable
Description: ocusec_2 is a categorical variable that specifies the sector of activity in the last 7 days. It classifies the secondary job's sector of activity of any individual with a job (lstatus=1) and is missing otherwise. The variable is constructed for all individuals that respond to this question, even if they are below the working age.
Four categories after harmonization:
1 = Public sector, Central Government, Army (including armed forces)
2 = Private, NGO; 3 = State-owned
4 = Public or State-owned, but cannot distinguish

Variable: industry_orig_2

Label: Sector of activity, secondary job (7-day ref period)
Type: String variable
Description: industry_orig_2 is a string variable that specifies the original industry codes for the second job with reference period of the last 7 days and should correspond to whatever is in the original file with

no recoding. Do not put missing values for people below the working age if they have a job. It classifies the main job of any individual with a job (lstatus=1) and is missing otherwise

Variable: industrycat10_2

Label: 1 digit industry classification, secondary job (7-day ref period)
Type: Numeric categorical variable
Description: industrycat10_2 is a categorical variable that specifies the 1-digit industry classification that classifies the second job with reference period of the last 7 days of any individual with a job (lstatus=1) and is missing otherwise. The variable is constructed for all individuals that respond to this question, even if they are below the working age. The codes for the second job are given here based on the UN International Standard Industrial Classification. Ten categories after harmonization:

1 = Agriculture, Hunting, Fishing, etc.	2 = Mining
3 = Manufacturing	4 = Public Utility Services
5 = Construction	6 = Commerce
7 = Transport and Communications	8 = Financial and Business Services
9 = Public Administration	10 = Other Services, Unspecified

Variable: industrycat4_2

Label: 4-category industry classification, secondary job (7-day ref period)
Type: Numeric categorical variable
Description: industrycat4_2 is a categorical variable that specifies the 1-digit industry classification for Broad Economic Activities for the second job with reference period of the last 7 days. This variable is either created directly from the data (if industry classification does not exist for 10 categories) or created from industrycat10_2.
Four categories after harmonization: 1 = Agriculture; 2= Industry; 3 = Services; 4 = Other

Variable: occup_orig_2

Label: Sector of activity, secondary job (7-day ref period)
Type: String variable
Description: occup_orig_2 is a string variable that specifies the original occupation code in the last 7 days for the secondary job. This variable corresponds to whatever is in the original file with no recoding.

Variable: occup_2

Label: 1 digit occupational classification, secondary job (7-day ref period)
Type: Numeric categorical variable
Description: occup_2 is a categorical variable that specifies the 1-digit occupation classification. It classifies the second job of any individual with a job (lstatus=1) and is missing otherwise. The variable is constructed for all individuals that respond to this question, even if they are below the working age. Most surveys collect detailed information and then code it, without keeping the original data. No attempt has been made to correct or check the original coding. The classification is based on the International Standard Classification of Occupations (ISCO). In the case of different classifications, re-coding has been done to best match the ISCO.
Eleven categories after harmonization:

1 = Managers	2 = Professionals
3 = Technicians and associate professionals	4 = Clerical support workers

5 = Service and sales workers
6 = Skilled agricultural, forestry and fishery workers
7 = Craft and related trades workers
8 = Plant and machine operators, and assemblers
9 = Elementary occupations
10 = Armed forces occupations
99 = Other/unspecified

Variable: wage_nc_2

Label: Last wage payment, secondary job, excl. bonuses, etc. (7-day ref period)
Type: Numeric continuous variable
Description: wage_nc_2 is a continuous variable that specifies the last wage payment in local currency of any individual (lstatus=1 & empstat_2<=4) in its secondary occupation and is missing otherwise. The wage should come from the second job, in other words, the job that the person dedicated the second most amount of time in the week preceding the survey. This excludes tips, bonuses, other compensation such as dwellings or clothes, and other payments. The variable is constructed for all persons administered this module in each questionnaire. For this reason, the lower age cutoff (and perhaps upper age cutoff) will vary from country to country. Notes:

- For all those with self-employment or owners of own businesses, this should be net revenues (net of all costs EXCEPT for tax payments) or the amount of salary taken from the business. Due to the almost complete lack of information on taxes, the wage from main job is NOT net of taxes.
- By definition, non-paid employees (empstat_2=2) should have wage=0.
- The reference period of the wage_nc_2 will be recorded in the unitwage_2 variable

Variable: unitwage_2

Label: Time unit of last wages payment, secondary job (7-day ref period)
Type: Numeric categorical variable
Description: unitwage_2 is a categorical variable that specifies the time reference for the wage_nc_2 variable. It specifies the time unit measurement for the wages for the secondary job of any individual (lstatus=1 & empstat_2=1) and is missing otherwise.
Ten categories after harmonization:
1 = Daily
2 = Weekly
3 = Every two weeks
4 = Every two months
5 = Monthly
6 = Quarterly
7 = Every six months
8 = Annually
9 = Hourly
10 = Other

Variable: whours_2

Label: Hours of work in last week, secondary job (7-day ref period)
Type: Numeric continuous variable
Description: whours_2 is a continuous variable that specifies the hours of work in last week for the second job with reference period of the last 7 days of any individual with a job (lstatus=1) and is missing otherwise. The second job defined as that occupation that the person dedicated the second most amount of time to over the past week. The variable is constructed for all persons administered this module in each questionnaire. The lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country. Notes:

- If the respondent was absent from the job in the week preceding the survey due to holidays, vacation, or sick leave, then record the time worked in the previous 7 days that the person worked.
- Sometimes the questions are phrased as, “on average, how many hours a week do you work?”.
- For individuals who only give information on how many hours they work per day and no information on number of days worked a week, multiply the hours by 5 days.

- In the case of a question that has hours worked per month, divide by 4.3 to get weekly hours.

Variable: wmonths_2
Label: Months worked in the last 12 months, secondary job (7-day ref period)
Type: Numeric continuous variable
Description: wmonths_2 is a continuous variable that specifies the number of months worked in the last 12 months for the secondary job of any individual with a job (lstatus=1) and is missing otherwise. The variable is constructed for all persons administered this module in each questionnaire.

Variable: wage_total_2
Label: Annualized total wage, secondary job (7-day ref period)
Type: Numeric continuous variable
Description: wage_total_2 is a continuous variable that specifies the annualized wage payment (regular wage plus bonuses, in-kind, compensation, etc.) in local currency of any individual (lstatus=1 & empstat_2=1) in its secondary occupation and is missing otherwise. The wage should come from the secondary job, in other words, the job that the person dedicated the second most amount of time in the week preceding the survey. This wage includes tips, compensations such as bonuses, dwellings or clothes, and other payments. wage_total_2 should be equal to wage_nc_2 in case there are no bonuses, tips etc. offered as part of the job. The variable is constructed for all persons administered this module in each questionnaire. For this reason, the lower age cutoff (and perhaps upper age cutoff) will vary from country to country. Notes:

- The annualization of the wage_total_2 should consider the number of months/weeks the persons have been working and receiving this income. You should not assume the respondent worked for the whole year.

Variable: firmsize_l_2
Label: Firm size (lower bracket), secondary job (7-day ref period)
Type: Numeric continuous variable
Description: firmsize_l_2 specifies the lower bracket of the firm size. The variable is constructed for all persons who are employed. If it is continuous, it records the number of people working for the same employer. If the variable is categorical, it records the lower boundary of the bracket.

Variable: firmsize_u_2
Label: Firm size (upper bracket), secondary job (7-day ref period)
Type: Numeric continuous variable
Description: firmsize_u_2 specifies the upper bracket of the firm size. The variable is constructed for all persons who are employed. If it is continuous, it records the number of people working for the same employer. If the variable is categorical, it records the upper boundary of the bracket. If the right bracket is open, a missing value should be inputted.

5.5 Other Employment, 7-day reference period

Variable: t_hours_others
Label: Annualized hours worked in all but primary and secondary jobs (7-day ref period)
Type: Numeric continuous variable

Description: t_hours_others is a continuous variable that specifies the hours of work in last 12 months in all jobs excluding the primary and secondary ones. If the respondent was absent from the job in the week preceding the survey due to holidays, vacation, or sick leave, then record the time worked in the previous 7 days that the person worked.

Variable: t_wage_nc_others
Label: Annualized wage in all but primary & secondary jobs excl. bonuses, etc. (7-day ref period)
Type: Numeric continuous variable
Description: t_wage_nc_others is a continuous variable that specifies the annualized wage in all jobs excluding the primary and secondary ones. This excludes tips, bonuses, other compensation such as dwellings or clothes, and other payments.

Variable: t_wage_others
Label: Annualized wage in all but primary and secondary jobs (7-day ref period)
Type: Numeric continuous variable
Description: t_wage_others is a continuous variable that specifies the annualized wage in all jobs excluding the primary and secondary ones. This wage includes tips, compensations such as bonuses, dwellings or clothes, and other payments. t_wage_others should be equal to t_wage_nc_ others in case there are no bonuses, tips etc. offered as part of any of the jobs.

5.6 Total Employment Earnings, 7-day reference period

Variable: t_hours_total
Label: Annualized hours worked in all jobs (7-day ref period)
Type: Numeric continuous variable
Description: t_hours_total is a continuous variable that specifies the hours of work in last 12 months in all jobs including primary, secondary and others. Note: if the respondent was absent from the job in the week preceding the survey due to holidays, vacation, or sick leave, then record the time worked in the previous 7 days that the person worked.

Variable: t_wage_nc_total
Label: Annualized wage in all jobs excl. bonuses, etc. (7-day ref period)
Type: Numeric continuous variable
Description: t_wage_nc_total is a continuous variable that specifies the total annualized wage income in all jobs including primary, secondary and others. This excludes tips, bonuses, other compensation such as dwellings or clothes, and other payments.

Variable: t_wage_total
Label: Annualized total wage for all jobs (7-day ref period)
Type: Numeric continuous variable
Description: t_wage_total is a continuous variable that specifies the total annualized wage income in all jobs including primary, secondary and others. This income includes tips, compensations such as bonuses, dwellings or clothes, and other payments. t_wage_total should be equal to t_wage_nc_total in case there are no bonuses, tips etc. offered as part of any of the jobs. If the number of months worked in this job is missing you could assumed that the person worked the whole year in this job.

5.7 Labor status, 12-month reference period

Variable: minlaborage_year

Label: Labor module application age (12-mon ref period)
Type: Numeric discrete variable
Description: This is the lowest age for which the labor module is implemented in the survey or the minimum working age in the country. For this reason, the lower age cutoff at which information is collected will vary from country to country.

Variable: lstatus_year

Label: Labor status (12-mon ref period)
Type: Numeric categorical variable
Description: lstatus_year is an individual's labor status in the last 12 months. The value must be missing for individuals less than the required age (minlaborage).
Three categories are used after harmonization:
1 = Employed; 2 = Unemployed; 3 = Not-in-labor force
All persons are considered active in the labor force if they presently have a job (formal or informal, i.e., employed) or do not have a job but are actively seeking work (i.e., unemployed).

Variable: nlfreason_year

Label: Reason not in the labor force (12-mon ref period)
Type: Numeric categorical variable
Description: nlfreason_year is the reason an individual was not in the labor force in the last 12 months. This variable is constructed for all those who are not presently employed and are not looking for work (lstatus_year=3) and missing otherwise.
Five categories after harmonization:
1= Student (a person currently studying.)
2= Housewife (a person who takes care of the house, older people, or children)
3= Retired
4 = Disabled (a person who cannot work due to physical conditions)
5 = Other (a person does not work for any other reason)
Fill this information for all people interviewed in the labor section of the questionnaire regardless of their age.

Variable: unempldur_l_year

Label: Unemployment duration (months) lower bracket (12-mon ref period)
Type: Numeric continuous variable
Description: unempldur_l_year is a continuous variable specifying the duration of unemployment in months (lower bracket).
The variable is constructed for all unemployed persons (lstatus_year=2, otherwise missing). If it is specified as continuous in the survey, it records the numbers of months in unemployment. If the variable is categorical it records the lower boundary of the bracket.

Variable: unempldur_u_year

Label: Unemployment duration (months) upper bracket (12-mon ref period)

Type: Numeric continuous variable

Description: unempldur_u_year is a continuous variable specifying the duration of unemployment in months (upper bracket). The variable is constructed for all unemployed persons (lstatus_year=2, otherwise missing). If it is specified as continuous in the survey, it records the numbers of months in unemployment. If the variable is categorical it records the upper boundary of the bracket. If the right bracket is open a missing value should be inputted. If the duration of unemployment is not reported as a range, but as continuous variables, the unempldur_l_year and unempldur_u_year variables will have the same value. If the high range is open-ended the unempldur_u_year variable will be missing.

5.8 Primary Employment, 12-month reference period

Variable: empstat_year

Label: Employment status, primary job (12-mon ref period)

Type: Numeric categorical varaible

Description: empstat is a categorical variable that specifies the main employment status in the last 12 months of any individual with a job (lstatus_year =1) and is missing otherwise. The variable is constructed for all individuals that respond to this question, even if they are below the working age. For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country.

The definitions are taken from the International Labor Organization's Classification of Status in Employment with some revisions to consider the data available. Five categories after harmonization:

1 = Paid Employee; 2 = Non-Paid Employee; 3 = Employer; 4 = Self-employed; 5 = Other, workers not classifiable by status

Variable: ocusec_year

Label: Sector of activity, primary job (12-mon ref period)

Type: Numeric categorical variable

Description: ocusec_year is a categorical variable that specifies the sector of activity in the last 12 months. It classifies the main job's sector of activity of any individual with a job (lstatus_year =1) and is missing otherwise. The variable is constructed for all individuals that respond to this question, even if they are below the working age.

Four categories after harmonization:

1 = Public sector, Central Government, Army (including armed forces)

2 = Private, NGO

3 = State-owned

4 = Public or State-owned, but cannot distinguish

Note: Do not code basis of occupation (ISCO) or industry (ISIC) codes.

Variable: industry_orig_year

Label: Original industry code, primary job (12-mon ref period)

Type: String variable

Description: industry_orig_year is a string variable that specifies the original industry codes in the last 12 months for the main job provided in the survey (the actual question) and should correspond to whatever

is in the original file with no recoding. It will contain missing values for people below the working age. It classifies the main job of any individual with a job (lstatus_year =1) and is missing otherwise

Variable: industrycat10_year

Label: 1 digit industry classification, secondary job (12-mon ref period)
Type: Numeric categorical variable
Description: industrycat10_year is a categorical variable that specifies the 1-digit industry classification in the last 12 months for the main job of any individual with a job (lstatus_year =1) and is missing otherwise. The variable is constructed for all individuals that respond to this question, even if they are below the working age. The codes for the main job are given here based on the UN International Standard Industrial Classification. It classifies the main job of any individual with a job (lstatus_year =1) and is missing otherwise.
Ten categories after harmonization:

1 = Agriculture, Hunting, Fishing, etc.	2 = Mining
3 = Manufacturing	4 = Public Utility Services
5 = Construction	6 = Commerce
7 = Transport and Communications	8 = Financial and Business Services
9 = Public Administration	10 = Other Services, Unspecified

Variable: industrycat4_year

Label: 4-category industry classification, secondary job (12-mon ref period)
Type: Numeric categorical variable
Description: industrycat4_year is a categorical variable that specifies the 1-digit industry classification in the last 12 months for the main job for Broad Economic Activities. This variable is either created directly from the data (if industry classification does not exist for ten categories) or created from industrycat10_year. Four categories after harmonization:
1 = Agriculture; 2= Industry; 3 = Services; 4 = Other
This variable is either created directly from the data (if industry classification does not exist for ten categories) or created from industrycat10_year.

Variable: occup_orig_year

Label: Original occupational classification, primary job (12-mon ref period)
Type: String variable
Description: occup_orig_year is a string variable that specifies the original occupation code in the last 12 months for the main job. This variable corresponds to whatever is in the original file with no recoding.

Variable: occup_year

Label: 1 digit occupational classification, primary job (12-mon ref period)
Type: Numeric categorical variable
Description: occup_year is a categorical variable that specifies the 1-digit occupational classification for the main job in the last 12 months of any individual with a job (lstatus_year =1) and is missing otherwise. The variable is constructed for all individuals that respond to this question, even if they are below the working age. For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country. The classification is based on the International Standard Classification of Occupations (ISCO). It classifies the main job of any individual with a job (lstatus_year=1) and is missing otherwise. Eleven categories after harmonization:

1 = Managers	2 = Professionals
3 = Technicians and associate professionals	4 = Clerical support workers

5 = Service and sales workers
6 = Skilled agricultural, forestry and fishery workers
7 = Craft and related trades workers
8 = Plant and machine operators, and assemblers
9 = Elementary occupations
10 = Armed forces occupations
99 = Other/unspecified

Variable: wage_nc_year

Label: Last wage payment, primary job, excl. bonuses, etc. (12-mon ref period)
Type: Numeric continuous variable
Description: wage_nc_year is a continuous variable that specifies the last wage payment in local currency of any individual (lstatus_year =1 & empstat_year =1) in its primary occupation at the reference period reported in the survey and it is missing otherwise. The wage should come from the main job, in other words, the job that the person dedicated most time in the 12 months preceding the survey. This excludes tips, bonuses, other compensation such as dwellings or clothes, and other payments. The variable is constructed for all persons administered this module in each questionnaire. For this reason, the lower age cutoff (and perhaps upper age cutoff) will vary from country to country. Notes:

- For all those with self-employment or owners of own businesses, this should be net revenues (net of all costs EXCEPT for tax payments) or the amount of salary taken from the business. Due to the almost complete lack of information on taxes, the wage from main job is NOT net of taxes.
- By definition, non-paid employees (empstat_year=2) should have wage=0.
- The reference period of the wage_nc_year will be recorded in the unitwage_year variable.

Variable: unitwage_year

Label: Time unit of last wages payment, primary job (12-mon ref period)
Type: Numeric categorical variable
Description:
unitwage_year is a categorical variable that specifies the time reference for the wage_nc_year variable. It specifies the time unit measurement for the wages of any individual (lstatus_year =1 & empstat_year =1) and it is missing otherwise. Acceptable values include:

1 = Daily
2 = Weekly
3 = Every two weeks
4 = Every two months
5 = Monthly
6 = Quarterly
7 = Every six months
8 = Annually
9 = Hourly
10 = Other

Variable: whours_year

Label: Hours of work in last week, primary job (12-mon ref period)
Type: Numeric continuous variable
Description: whours_year is a continuous variable that specifies the hours of work last week for the main job of any individual with a job (lstatus_year =1) and is missing otherwise. The main job defined as that occupation that the person dedicated more time to over the past 12 months. The variable is constructed for all persons administered this module in each questionnaire. Notes:

- Sometimes the questions are phrased as, “on average, how many hours a week do you work?”.
- For individuals who only give information on how many hours they work per day and no information on number of days worked a week, multiply the hours by 5 days.
- In the case of a question that has hours worked per month, divide by 4.3 to get weekly hours.

Variable: wmonths_year

Label: Months worked in the last 12 months, primary job (12-mon ref period)

Type: Numeric continuous variable

Description: wmonths_year is a continuous variable that specifies the number of months worked in the last 12 months for the main job of any individual with a job (lstatus_year =1) and is missing otherwise. The main job is defined as that occupation that the person dedicated more time to over the past 12 months. The variable is constructed for all persons administered this module in each questionnaire.

Variable: wage_total_year

Label: Annualized total wage, primary job (12-mon ref period)

Type: Numeric continuous variable

Description: wage_total_year is a continuous variable that specifies the annualized wage payment (regular wage plus bonuses, in-kind, compensation, etc.) for the primary occupation in local currency of any individual (lstatus_year =1 & empstat_year =1) and is missing otherwise. The wage should come from the main job, in other words, the job that the person dedicated most time in the year preceding the survey. This wage includes tips, compensations such as bonuses, dwellings or clothes, and other payments. wage_total_year should be equal to wage_nc_year in case there are no bonuses, tips etc. offered as part of the job. The variable is constructed for all persons administered this module in each questionnaire. The annualization of the wage_total_year should consider the number of months/weeks the persons have been working and receiving this income. You should not assume that the respondent worked for the whole year.

Variable: contract_year

Label: Contract (12-mon ref period)

Type: Numeric categorical variable

Description: contract_year is a dummy variable that classifies the contract status (yes/no) of any individual with a job (lstatus_year =1) and is missing otherwise. It indicates whether a person has a signed (formal) contract, regardless of duration. The variable is constructed for all persons administered this module in each questionnaire. Two categories after harmonization: 0 = No; 1 = Yes

Variable: healthins_year

Label: Health insurance (12-mon ref period)

Type: Numeric categorical variable

Description: healthins_year is a dummy variable that classifies the health insurance status (yes/no) of any individual with a job (lstatus_year =1) and is missing otherwise. Variable is constructed for all persons administered this module in each questionnaire. However, this variable is only constructed if there is an explicit question about health insurance provided by the job. Two categories after harmonization:
0 = No; 1 = Yes

Variable: socialsec_year

Label: Social security (12-mon ref period)

Type: Numeric categorical variable

Description: socialsec_year is a dummy variable that classifies the social security status (yes/no) of any individual with a job (lstatus_year =1) and is missing otherwise. Variable is constructed for all persons administered this module in each questionnaire. For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country. However, this variable is

only constructed if there is an explicit question about pension plans or social security. Two categories after harmonization: 0 = No; 1 = Yes

Variable: union_year

Label: Union membership (12-mon ref period)

Type: Numeric categorical variable

Description: union_year is a dummy variable that classifies the union membership status (yes/no) of any individual with a job (lstatus_year =1) and is missing otherwise. Variable is constructed for all persons administered this module in each questionnaire. For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country. However, this variable is only constructed if there is an explicit question about trade unions. Two categories after harmonization: 0 = No; 1 = Yes

Variable: firmsize_l_year

Label: Firm size (lower bracket), primary job (12-mon ref period)

Type: Numeric continuous variable

Description: firmsize_l_year specifies the lower bracket of the firm size. The variable is constructed for all persons who are employed in the last 12 months for the main job. If it is continuous, it records the number of people working for the same employer. If the variable is categorical, it records the lower boundary of the bracket.

Variable: firmsize_u_year

Label: Firm size (upper bracket), primary job (12-mon ref period)

Type: Numeric continuous variable

Description: firmsize_u_year specifies the upper bracket of the firm size. The variable is constructed for all persons who are employed in the last 12 months for the main job. If it is continuous, it records the number of people working for the same employer. If the variable is categorical, it records the upper boundary of the bracket. If the right bracket is open, this variable should be missing.

5.9 Secondary Employment, 12-month reference period

Variable: empstat_2_year

Label: Employment status, secondary job (12-mon ref period)

Type: Numeric categorical variable

Description: empstat_2_year is a categorical variable that specifies employment status of the secondary job with reference period of last 12 months of any individual with a job (lstatus_year =1) and is missing otherwise. The variable is constructed for all individuals that respond to this question, even if they are below the working age. For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country.

The definitions are taken from the International Labor Organization's Classification of Status in Employment with some revisions to consider the data available.

Five categories after harmonization:

1 = Paid Employee
2 = Non-Paid Employee
3 = Employer
4 = Self-employed
5 = Other, workers not classifiable by status

Variable: ocusec_2_year

Label: Sector of activity, secondary job (12-mon ref period)

Type: Numeric categorical variable

Description: ocusec_2_year is a categorical variable that specifies the sector of activity in the last 12 months. It classifies the secondary job's sector of activity of any individual with a job (lstatus_year =1) and is missing otherwise. The variable is constructed for all individuals that respond to this question, even if they are below the working age. Four categories after harmonization:
1 = Public sector, Central Government, Army (including armed forces)
2 = Private, NGO
3 = State-owned
4 = Public or State-owned, but cannot distinguish
Notes: Do not code basis of occupation (ISCO) or industry (ISIC) codes.

Variable: industry_orig_2_year

Label: Original industry code, secondary job (12-mon ref period)
Type: String variable
Description: industry_orig_2_year is a string variable that specifies the original industry codes for the second job with reference period of the last 12 months and should correspond to whatever is in the original file with no recoding. The variable is constructed for all individuals that respond to this question, even if they are below the working age. It classifies the main job of any individual with a job (lstatus_year=1) and is missing otherwise

Variable: industrycat10_2_year

Label: 1 digit industry classification, secondary job (12-mon ref period)
Type: Numeric categorical variable
Description: industrycat10_2_year is a categorical variable that specifies the 1-digit industry classification that classifies the second job with reference period of the last 12 months of any individual with a job (lstatus_year =1) and is missing otherwise. The variable is constructed for all individuals that respond to this question, even if they are below the working age. The codes for the second job are given here based on the UN International Standard Industrial Classification.
Ten categories after harmonization:

1 = Agriculture, Hunting, Fishing, etc.	2 = Mining
3 = Manufacturing	4 = Public Utility Services
5 = Construction	6 = Commerce
7 = Transport and Communications	8 = Financial and Business Services
9 = Public Administration	10 = Other Services, Unspecified

Variable: industrycat4_2_year

Label: 4-category industry classification, secondary job (12-mon ref period)
Type: Numeric categorical variable
Description: industrycat4_2_year is a categorical variable that specifies the 1-digit industry classification for Broad Economic Activities for the second job with reference period of the last 12 months. This variable is either created directly from the data (if industry classification does not exist for 10 categories) or created from industrycat10_year. Four categories after harmonization:
1 = Agriculture; 2= Industry; 3 = Services; 4 = Other
This variable is either created directly from the data (if industry classification does not exist for 10 categories) or created from industrycat10_2_year.

Variable: occup_orig_2_year

Label: Original occupational classification, secondary job (12-mon ref period)
Type: String variable

Description: occup_orig_2_year is a string variable that specifies the original occupation code in the last 12 months for the secondary job. This variable corresponds to the original file with no recoding.

Variable: occup_2_year

Label: 1 digit occupational classification, secondary job (12-mon ref period)
Type: Numeric categorical variable
Description: occup_2_year is a categorical variable that specifies the 1-digit occupation classification. It classifies the second job of any individual with a job (lstatus_year =1) and is missing otherwise. The variable is constructed for all individuals that respond to this question, even if they are below the working age. Most surveys collect detailed information and then code it, without keeping the original data. No attempt has been made to correct or check the original coding. The classification is based on the International Standard Classification of Occupations (ISCO). In the case of different classifications, re-coding has been done to best match the ISCO.
Eleven categories after harmonization:

1 = Managers	2 = Professionals
3 = Technicians and associate professionals	4 = Clerical support workers
5 = Service and sales workers	6 = Skilled agricultural, forestry and fishery workers
7 = Craft and related trades workers	8 = Plant and machine operators, and assemblers
9 = Elementary occupations	10 = Armed forces occupations
99 = Other/unspecified	

Variable: wage_nc_2_year

Label: Last wage payment, secondary job, excl. bonuses, etc. (12-mon ref period)
Type: Numeric continuous variable
Description: wage_nc_2_year is a continuous variable that specifies the last wage payment in local currency of any individual (lstatus_year =1 & empstat_2_year =1) in its secondary occupation and is missing otherwise. The wage should come from the second job, in other words, the job that the person dedicated the second most amount of time in the week preceding the survey. This excludes tips, bonuses, other compensation such as dwellings or clothes, and other payments. The variable is constructed for all persons administered this module in each questionnaire. For this reason, the lower age cutoff (and perhaps upper age cutoff) will vary from country to country. Notes:

- For all those with self-employment or owners of own businesses, this should be net revenues (net of all costs EXCEPT for tax payments) or the amount of salary taken from the business. Due to the almost complete lack of information on taxes, the wage from main job is NOT net of taxes.
- By definition, non-paid employees (empstat_year_2 =2) should have wage=0.
- The reference period of the wage_nc_year_2 will be in the unitwage_2_year variable

Variable: unitwage_2_year

Label: Time unit of last wages payment, secondary job (12-mon ref period)
Type: Numeric categorical variable
Description: unitwage_2_year is a categorical variable that specifies the time reference for the wage_nc_2_year variable. It specifies the time unit measurement for the wages for the secondary job of any individual (lstatus_year =1 & empstat_2_year =1) and is missing otherwise.
Ten categories after harmonization:

1 = Daily	2 = Weekly
3 = Every two weeks	4 = Every two months
5 = Monthly	6 = Quarterly
7 = Every six months	8 = Annually

9 = Hourly 10 = Other

Variable: whours_2_year

Label: Hours of work in last week, secondary job (12-mon ref period)

Type: Numeric continuous variable

Description: whours_2_year is a continuous variable that specifies the hours of work in last week for the second job with reference period of the last 12 months of any individual with a job (lstatus_year =1) and is missing otherwise. The second job defined as that occupation that the person dedicated the second most amount of time to over the past year. The variable is constructed for all persons administered this module in each questionnaire. The lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country. Notes:

- Sometimes the questions are phrased as, "on average, how many hours a week do you work?".
- For individuals who only give information on how many hours they work per day and no information on number of days worked a week, multiply the hours by 5 days.
- In the case of a question that has hours worked per month, divide by 4.3 to get weekly hours.

Variable: wmonths_2_year

Label: Months worked in the last 12 months, secondary job (12-mon ref period)

Type: Numeric continuous variable

Description: wmonths_2_year is a continuous variable that specifies the number of months worked in the last 12 months for the secondary job of any individual with a job (lstatus_year =1) and is missing otherwise. The variable is constructed for all persons administered this module in each questionnaire.

Variable: wage_total_2_year

Label: Annualized total wage, secondary job (12-mon ref period)

Type: Numeric continuous variable

Description: wage_total_2_year is a continuous variable that specifies the annualized wage payment (regular wage plus bonuses, in-kind, compensation, etc.) in local currency of any individual (lstatus_year =1 & empstat_2_year =1) in its secondary occupation and is missing otherwise. The wage should come from the secondary job, in other words, the job that the person dedicated the second most amount of time in the year preceding the survey. This wage includes tips, compensations such as bonuses, dwellings or clothes, and other payments. wage_total_2_year should be equal to wage_nc_2_year in case there are no bonuses, tips etc. offered as part of the job. The variable is constructed for all persons administered this module in each questionnaire. For this reason, the lower age cutoff (and perhaps upper age cutoff) will vary from country to country. Notes:

- The annualization of the wage_total_2_year should consider the number of months/weeks the persons have been working and receiving this income. You should not assume that the respondent worked for the whole year.

Variable: firmsize_l_2_year

Label: Firm size (lower bracket), secondary job (12-mon ref period)

Type: Numeric continuous variable

Description: firmsize_l_2_year specifies the lower bracket of the firm size. The variable is constructed for all persons who are employed. If it is continuous, it records the number of people working for the same employer. If the variable is categorical, it records the lower boundary of the bracket.

Variable: firmsize_u_2_year

Label: Firm size (upper bracket), secondary job (12-mon ref period)
Type: Numeric continuous variable
Description: firmsize_u_2_year specifies the upper bracket of the firm size. The variable is constructed for all persons who are employed. If it is continuous, it records the number of people working for the same employer. If the variable is categorical, it records the upper boundary of the bracket. If the right bracket is open, a missing value should be inputted.

5.10 Other Employment, 12-month reference period

Variable: t_hours_others_year

Label: Annualized hours worked in all but primary and secondary jobs (12-mon ref period)
Type: Numeric continuous variable
Description: t_hours_others_year is a continuous variable that specifies the hours of work in last 12 months in all jobs excluding the primary and secondary ones.

Variable: t_wage_nc_others_year

Label: Annualized wage in all but primary & secondary jobs excl. bonuses, etc. (12-mon ref period)
Type: Numeric continuous variable
Description: t_wage_nc_others_year is a continuous variable that specifies the annualized wage in last 12 months in all jobs excluding the primary and secondary ones. This excludes tips, bonuses, other compensation such as dwellings or clothes, and other payments.

Variable: t_wage_others_year

Label: Annualized wage in all but primary and secondary jobs (12-mon ref period)
Type: Numeric continuous variable
Description: t_wage_others_year is a continuous variable that specifies the annualized wage in last 12 months in all jobs excluding the primary and secondary ones. This wage includes tips, compensations such as bonuses, dwellings or clothes, and other payments. t_wage_others should be equal to t_wage_nc_ others in case there are no bonuses, tips etc. offered as part of any of the jobs.

5.11 Total Employment Earnings, 12-month reference period

Variable: t_hours_total_year

Label: Annualized hours worked in all jobs (12-mon ref period)
Type: Numeric continuous variable
Description: t_hours_total_year is a continuous variable that specifies the hours of work in last 12 months in all jobs including primary, secondary and others.

Variable: t_wage_nc_total_year

Label: Annualized wage in all jobs excl. bonuses, etc. (12-mon ref period)
Type: Numeric continuous variable
Description: t_wage_nc_total_year is a continuous variable that specifies the total annualized wage income in all jobs including primary, secondary and others. This excludes tips, bonuses, other compensation such as dwellings or clothes, and other payments.

Variable: t_wage_total_year

Label: Annualized total wage for all jobs (12-mon ref period)

Type: Numeric continuous variable

Description: t_wage_total_year is a continuous variable that specifies the total annualized wage income in all jobs including primary, secondary and others. This income includes tips, compensations such as bonuses, dwellings or clothes, and other payments. t_wage_total_year should be equal to t_wage_nc_total in case there are no bonuses, tips etc. offered as part of any of the jobs.

5.12 Total Labor Income

Variable: njobs

Label: Total number of jobs

Type: Numeric continuous variable

Description: njobs is a numeric variable that specifies the total number of jobs. Do not put missing value for people below working age, unemployed and people out of the labor force.

Variable: t_hours_annual

Label: Total hours worked in all jobs in the previous 12 months

Type: Numeric continuous variable

Description: t_hours_annual is a continuous variable that specifies the annual numbers of hours worked in all the jobs including primary, secondary and others regardless of their period of reference.

Variable: linc_nc

Label: Total annual wage income in all jobs, excl. bonuses, etc.

Type: Numeric continuous variable

Description: linc_nc is a continuous variable that specifies the total annualized wage income in all the jobs including primary, secondary and others regardless of their period of reference. This excludes tips, bonuses, other compensation such as dwellings or clothes, and other payments.

Variable: laborincome

Label: Total annual individual labor income in all jobs, incl. bonuses, etc.

Type: Numeric continuous variable

Description: laborincome is a continuous variable that specifies the total annualized individual labor income in all jobs including primary, secondary and others regardless of their period of reference. This income includes tips, compensations such as bonuses, dwellings or clothes, and other payments. This variable should be used as the total annual labor income of an individual.